

QY 714 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGACAGAGTTCAGAAACATAGAAACCA 773
 DB 121 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGACAGAGTTCAGAAACATAGAAACCA 180
 QY 774 ACTAATCAAGAAATATTTCTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTG 833
 DB 181 ACTAATCAAGAAATATTTCTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTG 240
 QY 834 TTTTGGGCGCAACAGAAACAAGACTCTGGTTAGCCATAAAAGATTTTACCCCT 893
 DB 241 TTTTGGGCGCAACAGAAACAAGACTCTGGTTAGCCATAAAAGATTTTACCCCT 300
 QY 894 TCAACCAATTTGGCACTAAAGAAATTTCTGTAGCTCTGCAATTTTGAATGACG 953
 DB 301 TCAACCAATTTGGCACTAAAGAAATTTCTGTAGCTCTGCAATTTTGAATGACG 360
 QY 954 TGAATGTGCAAAACAGTCTATTTGTTTATATTTGAATTTGTTTACCTATGA 1013
 DB 361 TGAATGTGCAAAACAGTCTATTTGTTTATATTTGAATTTGTTTACCTATGA 420
 QY 1014 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACAC 1073
 DB 421 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACAC 480
 QY 1074 TCTCTGGTTTA 1084
 DB 481 TCTCTGGTTTA 491
 RESULT 9
 AAI50961 standard; DNA; 494 BP.
 AC AAI50961;
 XX
 DT 17-OCT-2001 (first entry)
 DE Probe #19647 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KM genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001MO-US000663.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR,
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human placenta.
 PS Claim 25; SEQ ID NO 19647; 654bp; English.
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders
 XX
 SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
 Query Match 45.3%; Score 491; DB 4; Length 494;
 Best Local Similarity 100.0%; Pred. No. 2,4e-117;
 Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 594 TGGCAAGTGGGTGAACAAGACCAATGAACAGAAATTAATATAGACATGAAATGTA 653
 DB 1 TGGCAAGTGGGTGAACAAGACCAATGAACAGAAATTAATATAGACATGAAATGTA 60
 QY 654 AAGCCAGCCAGAAAAGGGGCGAGAGACATGTTGATTCCTACAGCTGTTCCAAATTTG 713
 DB 61 AAGCCAGCCAGAAAAGGGGCGAGAGACATGTTGATTCCTACAGCTGTTCCAAATTTG 120
 QY 714 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGACAGAGTTCAGAAACATAGAAACCA 773
 DB 121 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGACAGAGTTCAGAAACATAGAAACCA 180
 QY 774 ACTAATCAAGAAATATTTCTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTG 833
 DB 181 ACTAATCAAGAAATATTTCTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTG 240
 QY 834 TTTTGGGCGCAACAGAAACAAGACTCTGGTTAGCCATAAAAGATTTTACCCCT 893
 DB 241 TTTTGGGCGCAACAGAAACAAGACTCTGGTTAGCCATAAAAGATTTTACCCCT 300
 QY 894 TCAACCAATTTGGCACTAAAGAAATTTCTGTAGCTCTGCAATTTTGAATGACG 953
 DB 301 TCAACCAATTTGGCACTAAAGAAATTTCTGTAGCTCTGCAATTTTGAATGACG 360
 QY 954 TGAATGTGCAAAACAGTCTATTTGTTTATATTTGAATTTGTTTACCTATGA 1013
 DB 361 TGAATGTGCAAAACAGTCTATTTGTTTATATTTGAATTTGTTTACCTATGA 420
 QY 1014 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACAC 1073
 DB 421 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACAC 480
 QY 1074 TCTCTGGTTTA 1084
 DB 481 TCTCTGGTTTA 491
 RESULT 10
 ABA37272 standard; DNA; 494 BP.
 ID ABA37272
 XX
 AC ABA37272;
 XX
 DT 23-JAN-2002 (first entry)
 DE Probe #15738 for gene expression analysis in human heart cell sample.
 XX
 KW Human; gene expression; heart; microarray; vascular system; probe;
 KM cardiovascular disease; hypertension; cardiac arrhythmia;
 XX congenital heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157274-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001MO-US000666.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX
PS Claim 4; SEQ ID NO 15738; 530bp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WPI at ftp.wpi.edu/pub/published_pct_sequences
XX
XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
XX
XX Query Match 45.3%; Score 491; DB 4; Length 494;
XX Best Local Similarity 100.0%; Pred. No. 2,4e-117;
XX Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 594 TGGCAAAAGTGGGTGAACAGACATGAAACAGAAATTATTATGACATGATGTAA 653
DB 1 TGGCAAAAGTGGGTGAACAGACATGAAACAGAAATTATTATGACATGATGTAA 60
QY 654 AAGCCAGCCCAAGAAAGGGGCGAGAGACATGTTGATTCCTACGACTGTTCCAAATTG 713
DB 61 AAGCCAGCCCAAGAAAGGGGCGAGAGACATGTTGATTCCTACGACTGTTCCAAATTG 120
QY 714 TGTTAAGACCTTTAAACAAGTTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCA 773
DB 121 TGTTAAGACCTTTAAACAAGTTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCA 180
QY 774 ACTATACAGAAATATTTCTTACAGTGGAGAACCTTATCTGCGAAATGAAACATCTG 813
DB 181 ACTATACAGAAATATTTCTTACAGTGGAGAACCTTATCTGCGAAATGAAACATCTG 240
QY 813 TTTTGGGCGCAACAGAAACAGACTCTTGGTTAGCCATAAAGATTTTATTAACCCCT 893
DB 241 TTTTGGGCGCAACAGAAACAGACTCTTGGTTAGCCATAAAGATTTTATTAACCCCT 300
QY 894 TCAAACCACTTTGGCACTAAGAAATTCCTGTAAGTCTCTTGCAAATTTTGTATGAG 953
DB 301 TCAAACCACTTTGGCACTAAGAAATTCCTGTAAGTCTCTTGCAAATTTTGTATGAG 360
QY 954 TGAATGTCACAAACAGTTCTAATTTGTTTAAATTTGAAATTTGGTTTACCTATGA 1013
DB 361 TGAATGTCACAAACAGTTCTAATTTGTTTAAATTTGAAATTTGGTTTACCTATGA 420
QY 1014 AATTCCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1073
DB 421 AATTCCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
QY 1074 TCTCTGTTTA 1084
DB 481 TCTCTGTTTA 491
XX
XX RESULT 11
XX AAK44992
XX ID AAK44992 standard; DNA; 494 BP.
XX

AC AAK44992;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 19549.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000668.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 19549; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
XX
XX Query Match 45.3%; Score 491; DB 4; Length 494;
XX Best Local Similarity 100.0%; Pred. No. 2,4e-117;
XX Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 594 TGGCAAAAGTGGGTGAACAGACATGAAACAGAAATTATTATGACATGATGTAA 653
DB 1 TGGCAAAAGTGGGTGAACAGACATGAAACAGAAATTATTATGACATGATGTAA 60
QY 654 AAGCCAGCCCAAGAAAGGGGCGAGAGACATGTTGATTCCTACGACTGTTCCAAATTG 713
DB 61 AAGCCAGCCCAAGAAAGGGGCGAGAGACATGTTGATTCCTACGACTGTTCCAAATTG 120
QY 714 TGTTAAGACCTTTAAACAAGTTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCA 773
DB 121 TGTTAAGACCTTTAAACAAGTTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCA 180
QY 774 ACTATACAGAAATATTTCTTACAGTGGAGAACCTTATCTGCGAAATGAAACATCTG 813
DB 181 ACTATACAGAAATATTTCTTACAGTGGAGAACCTTATCTGCGAAATGAAACATCTG 240
QY 813 TTTTGGGCGCAACAGAAACAGACTCTTGGTTAGCCATAAAGATTTTATTAACCCCT 893
DB 241 TTTTGGGCGCAACAGAAACAGACTCTTGGTTAGCCATAAAGATTTTATTAACCCCT 300
QY 894 TCAAACCACTTTGGCACTAAGAAATTCCTGTAAGTCTCTTGCAAATTTTGTATGAG 953
DB 301 TCAAACCACTTTGGCACTAAGAAATTCCTGTAAGTCTCTTGCAAATTTTGTATGAG 360

CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS1005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 494 BP, 170 A, 82 C, 90 G, 152 T, 0 U, 0 Other;
SQ

Query Match 45.3%; Score 491; DB 4; Length 494;
Best Local Similarity 100.0%; Pred. No. 2,4e-117;
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 594 TGGCAAGTGGGTGAAACAGACATGAAACAGAAATTTATTATGACATGAAATGTA 653
DB 1 TGGCAAGTGGGTGAAACAGACATGAAACAGAAATTTATTATGACATGAAATGTA 60
DY 654 AAGCAGCCGAGAAAGGGGGGAGAGACATGTTGATCTTACGACTGTTCCAAATTTG 7.3
DB 61 AAGCAGCCGAGAAAGGGGGGAGAGACATGTTGATCTTACGACTGTTCCAAATTTG 120
DY 714 TGTTAAGACCTTTAACAAGTTGGCTGAATTTGAGCAGAGTTGACAGAAATAGAACCA 773
DB 121 TGTTAAGACCTTTAACAAGTTGGCTGAATTTGAGCAGAGTTGACAGAAATAGAACCA 180
DY 774 ACTATACAGAAATATTTCTTACAGTGAACACCTACTTATCTGGGAAATGAAACATCTG 833
DB 181 ACTATACAGAAATATTTCTTACAGTGAACACCTACTTATCTGGGAAATGAAACATCTG 240
DY 834 TTTTGGGCGAACAAGAAACAGACTCTGTTAGCCATAAAGATTTATTAACCCCT 893
DB 241 TTTTGGGCGAACAAGAAACAGACTCTGTTAGCCATAAAGATTTATTAACCCCT 300
DY 894 TCAACACGATTTGCCACATGAAGATTTCTGAGTCTCTCCAAATTTTGAATGAG 953
DB 301 TCAACACGATTTGCCACATGAAGATTTCTGAGTCTCTCCAAATTTTGAATGAG 360
DY 954 TGAATTTGCAAAACAGTTCTATTGTTTATTAATTTGAATTTGGTTTATTAACCTATGA 1013
DB 361 TGAATTTGCAAAACAGTTCTATTGTTTATTAATTTGAATTTGGTTTATTAACCTATGA 420
DY 1014 AATTCCTTTTAAATTAACATATGAAGAAATCCCTTACCTATCGAAGAAACAC 1073
DB 421 AATTCCTTTTAAATTAACATATGAAGAAATCCCTTACCTATCGAAGAAACAC 480
DY 1074 TCTCTGGTTTA 1084
DB 481 TCTCTGGTTTA 491

RESULT 14
ABS19237
ID ABS19237 standard; DNA; 494 BP.
XX ABS19237;
AC
XX
XX 19-AUG-2002 (first entry)
DE Human genome-derived single exon probe ORF from lung SEQ ID No 19228.
XX
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karagen syndrome;
KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KM primary ciliary dyskinesia; pulmonary hypertension;
KM hyaline membrane disease; open reading frame; ORF.
OS Homo sapiens.
XX WO200186003-A2.
XX
XX 15-NOV-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000665.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
XX Claim 4; SEQ ID NO 19228; 634p; English.

CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a single exon probe open reading frame of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 494 BP, 170 A, 82 C, 90 G, 152 T, 0 U, 0 Other;

Query Match 45.3%; Score 491; DB 6; Length 494;
 Best Local Similarity 100.0%; Pred. No. 2.4e-117;
 Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 594 TGGCAAGTGGTGAACAGCAATGAAACAGAAATTTATATAGACATGGATGTGA 653
 DB 1 TGGCAAGTGGTGAACAGCAATGAAACAGAAATTTATATAGACATGGATGTGA 60
 QY 654 AAGCCAGCCAGAAAAGGGGCGAGAGACATGGTTGATCTCAGACATGTTCCAAATTTG 713
 DB 61 AAGCCAGCCAGAAAAGGGGCGAGAGACATGGTTGATCTCAGACATGTTCCAAATTTG 120
 QY 714 TGTAAAGACCTTTAACAAGTTGGCTGAATTTGAGCAGAGTTCAAGAAATGAAACCA 773
 DB 121 TGTAAAGACCTTTAACAAGTTGGCTGAATTTGAGCAGAGTTCAAGAAATGAAACCA 190
 QY 774 ACTATACAGAAATTTCTTTTCACTGAGAGACCTATATGTGGAAATGAAACATG 833
 DB 181 ACTATACAGAAATTTCTTTTCACTGAGAGACCTATATGTGGAAATGAAACATG 240
 QY 834 TTTTGGGCGAACAAGAACTCTGTGGTTAGCATAAAGATTTTATACCCCT 893
 DB 241 TTTTGGGCGAACAAGAACTCTGTGGTTAGCATAAAGATTTTATACCCCT 300
 QY 894 TCAAAACACATTTGCCAATAAGAAATTTCTGTAGTCTCTTGCATAATTTTGATGAG 953
 DB 301 TCAAAACACATTTGCCAATAAGAAATTTCTGTAGTCTCTTGCATAATTTTGATGAG 360
 QY 954 TCAATGTGCAACAAGTCTATTTGTTTATATATTTGAATTTGTTTACCTATG 1013
 DB 361 TCAATGTGCAACAAGTCTATTTGTTTATATATTTGAATTTGTTTACCTATG 420
 QY 1014 AATTCCTTTTATTAATAATACATATGAGAAATCCCTTACCTACGAAACAAACAC 1073
 DB 421 AATTCCTTTTATTAATAATACATATGAGAAATCCCTTACCTACGAAACAAACAC 480
 QY 1074 TCTCTGTTTA 1084
 DB 481 TCTCTGTTTA 491

RESULT 15
 AAV88347
 ID AAV88347 standard; cDNA; 506 BP.
 XX
 AC AAV88347;
 XX
 DT 12-FEB-1999 (first entry)
 XX
 DE EST clone GB814.
 XX
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemokines; chemokines; haemostasis; gene therapy; thrombolytic;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 OS Homo sapiens.
 XX
 PN MO9845437-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US006956.
 XX
 PR 10-APR-1997; 97US-00837312.
 XX
 PA (GENY) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Werberg D, Treacy M;
 PI Spaulding V, Agoestino MJ;
 XX
 DR WPI; 1999-070078/06.
 XX

PT New polynucleotides encoding human secreted proteins - derived from e.g.
 PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,
 PT pituitary, retina and colon cDNA libraries.
 XX

BS Claim 1; Page 366; 641pp; English.

CC The present sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene therapy
 XX

SQ Sequence 506 BP; 171 A; 86 C; 119 G; 130 T; 0 U; 0 Other;

Query Match 37.0%; Score 401.4; DB 2; Length 506;
 Best Local Similarity 99.8%; Pred. No. 4.7e-94;
 Matches 402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 497 GCTGCTGCTTTTGAAGGAATGATGATGTTCACTGGAAGAAATGGAACATTAATT 556
 DB 99 GCTGCTGCTTTTGAAGGAATGATGATGTTCACTGGAAGAAATGGAACATTAATT 158
 QY 557 CAAGTAGCACTATATACGAAACATGTTCAACCAATGGCAAGTGGTGAACAGAG 616
 DB 159 CAAGTAGCACTATATACGAAACATGTTCAACCAATGGCAAGTGGTGAACAGAG 218
 QY 617 AATGAACAGAAATTTATATGAGACATGATGAAAGGCGCCGAAAGGGGGCA 676
 DB 219 AATGAACAGAAATTTATATGAGACATGATGAAAGGCGCCGAAAGGGGGCA 278
 QY 677 GAGACATGTTGATTCCTACAGCTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 736
 DB 279 GAGACATGTTGATTCCTACAGCTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 338
 QY 737 GCTGAATTTGAGCAGAGTTTAAGAACATAGAAACCAACTATACAAAGATTTCTTTAC 796
 DB 339 GCTGAATTTGAGCAGAGTTTAAGAACATAGAAACCAACTATACAAAGATTTCTTTAC 398
 QY 797 AGTGAGAACCTACTATCTGGGAAATGAAACATCTGTTTGGGCCAAGAGAAACAG 856
 DB 399 AGTGAGAACCTACTATCTGGGAAATGAAACATCTGTTTGGGCCAAGAGAAACAG 458
 QY 857 ACTCTGTTTAGCCATTAAGAAATTTTATTAACCCCTTCAAC 899
 DB 459 ACTCTGTTTAGCCATTAAGAAATTTTATTAACCCCTTCAAC 501

Search completed: April 22, 2004, 00:39:08
 Job time : 487.252 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 00:16:47 ; Search time 4418.69 Seconds

(Without alignments)
10632.969 Million cell updates/sec

Title: US-10-010-050A-1_COPY_1_1084

Sequence: 1 gaattcgccacgagggcaggg.....acaaacactctctgttca 1084

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sta:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sta:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rdd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1084	100.0	1486	6	BD073402
2	1067.8	98.5	4080	9	AF068227
3	1067	98.4	1751	6	BD136338
4	1064.6	98.2	2120	6	BD127858
5	1064.6	98.2	2120	9	AK075109
6	772.4	71.3	1038	6	BD073413
7	676.6	62.4	2318	10	BC025487
8	662.8	61.1	198524	2	AC109554
9	635.8	58.7	227920	2	AC107510
10	567	52.3	697	6	BD125187
11	567	52.3	697	6	BD126453
12	512.2	47.3	106988	9	AC001226
13	512.2	47.3	169362	2	AL136440
14	512.2	47.3	251187	2	AL1359875
15	401.4	37.0	506	6	BD060465
16	377.6	34.8	110000	10	AE014175_0
17	377.6	34.8	156823	2	AC102815
18	361.8	33.4	229015	2	AC106966
19	360.2	33.2	217336	2	AC113134
20	295.4	27.3	198524	2	AC109554
21	238.2	22.0	251187	2	AL1359875
22	199	18.4	258273	2	AC113855
23	133	12.3	473	6	AX333881
24	133	12.3	473	6	AX335993
25	55.2	5.1	110000	3	AC116305_0
26	51.4	4.7	313050	3	PFA929352
27	50.8	4.7	210095	2	AC135509
28	50.6	4.7	7218	6	166494
29	50.6	4.7	190664	9	AC090984
30	49.8	4.6	198718	2	AC136736
31	49.8	4.6	235063	2	AC107818
32	49.6	4.6	3535	1	AF125274
33	49.6	4.6	3535	1	AF125274
34	49.4	4.6	202347	2	BX663507
35	48.8	4.5	4490	9	MEFA012216
36	48.8	4.5	10742	1	AE010899
37	48.6	4.5	114639	9	AC100810
38	48.2	4.4	125020	9	AF429315
39	48.2	4.4	250029	3	AE014830
40	47.4	4.4	318221	2	PFMAL13P3
41	47	4.3	136996	10	AC116724
42	46.8	4.3	2000	6	AX655393
43	46.8	4.3	134971	2	AC116367
44	46.8	4.3	153751	3	AC116551
45	46.6	4.3	273222	2	BX572646

ALIGNMENTS

```
RESULT 1
LOCUS      BD073402                      1486 bp    DNA    linear    PAT 27-AUG-2002
DEFINITION Secreted protein which human chromosome 13 encodes.
ACCESSION  BD073402
VERSION    BD073402.1 GI:22619005
KEYWORDS   JP 2001511345-A/1.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE  1 (bases 1 to 1486)
AUTHORS   Sheppard,P.O. and Gilberton,D.G.
TITLE      Secreted protein which human chromosome 13 encodes
JOURNAL    Patent: JP 2001511345-A 1 14-AUG-2001,
```

COMMENT	ZYMOGENETICS INC	OS	Homo sapiens (human)
PN	JP 2001511345-A/1		
PD	14-AUG-2001		
PF	24-JUL-1998 JP 2000504249		
PI	24-JUL-1997 US 60/053613		
PI	PAUL O SHEPPARD, DIBRA G GILBERTON		
PC	C12N15/09, A61K48/00, A61K48/00, C07K14/47, C07K16/18, C12N1/15, PC C12N1/19,		
PC	C12N1/21, C12N5/10, C12P21/02, C12Q1/68, C12N15/00, A61K37/02, C12N5/00		
CC	Secreted protein which human chromosome 13 encodes	FH	Key
FT	Location/Qualifiers (47) . . (1084) .		
FEATURES	source		
	1. 1486		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	100.0%; Score 1084; DB 6; Length 1486;		
Best Local Similarity	100.0%; Pred. No. 1,1e-238;		
Matches 1084;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 GAATTGGCAGAGGGCGAGAGGTAGACACGGCACAGGGCGCCGAGATCGCGGGCGCG	60	
Db	1 GAATTGGCAGAGGGCGAGAGGTAGACACGGCACAGGGCGCCGAGATCGCGGGCGCG	60	
QY	GGGCGCGGCTGGGGAGCGGCTTCTCGTGCCTGGGCGCTGGCGGCTGCTTGGCTGGCGGT	120	
Db	GGGCGCGGCTGGGGAGCGGCTTCTCGTGCCTGGGCGGCTGGCGGCTGCTTGGCTGGCGGT	120	
QY	GATTCGGGGCTGGTCCCGGGCTCGGGGATCCCTCCCGGCGCCACTGGGCGGTCCTTA	180	
Db	GATTCGGGGCTGGTCCCGGGCTCGGGGATCCCTCCCGGCGCCACTGGGCGGTCCTTA	180	
QY	CAAGGCTTTGACTCCGTCGCAAAACCTGATCTTATTTGCAAGCTAAGTACTCTTCTG	240	
Db	CAAGGCTTTGACTCCGTCGCAAAACCTGATCTTATTTGCAAGCTAAGTACTCTTCTG	240	
QY	TCCAACTGGCTCACTATTCCTAGTTATGGAGGGTATGATGACATTGAATTTTGCATT	300	
Db	TCCAACTGGCTCACTATTCCTAGTTATGGAGGGTATGATGACATTGAATTTTGCATT	300	
QY	ACAAGCCCCAGTATGGAAATTTAAATATGAGAGCCTCCGCGGACACTTGAATAATGACA	360	
Db	ACAAGCCCCAGTATGGAAATTTAAATATGAGAGCCTCCGCGGACACTTGAATAATGACA	360	
QY	TGATGCCATTGGATTGAGAAGTACATTAACTGGCAGAACTACACAAATGGAATGGTATGA	420	
Db	TGATGCCATTGGATTGAGAAGTACATTAACTGGCAGAACTACACAAATGGAATGGTATGA	420	
QY	ACTTTTCCAACTTGGCAACTGATCACTTCCCATCTCCGACCTGAAATGGAATGCCCTTT	480	
Db	ACTTTTCCAACTTGGCAACTGATCACTTCCCATCTCCGACCTGAAATGGAATGCCCTTT	480	
QY	CTGGTGATATCAAGGCGGTGCTGCTTTTGAAGGAAATGATGATGTCCTGGAGAGA	540	
Db	CTGGTGATATCAAGGCGGTGCTGCTTTTGAAGGAAATGATGATGTCCTGGAGAGA	540	
QY	AAATGGGACATTAGTTCAAGTAGCAACTATATCAGAGAAACATGTCCAACCAATGGCAAA	600	
Db	AAATGGGACATTAGTTCAAGTAGCAACTATATCAGAGAAACATGTCCAACCAATGGCAAA	600	
QY	GTGGGTGAACAGGACCAATGAAACAGGAATTTATTTATGACATGGAATGTAAAAGCCAG	660	
Db	GTGGGTGAACAGGACCAATGAAACAGGAATTTATTTATGACATGGAATGTAAAAGCCAG	660	
QY	CCGAGAAAAGGGGGAGAGACATGGTTGATTTCCAGACCTGTTCCAATTTGCTTAG	720	
Db	CCGAGAAAAGGGGGAGAGACATGGTTGATTTCCAGACCTGTTCCAATTTGCTTTAAG	720	

QY	721	GACCTTTAACAGTGGCTGAATTTGGAGCAGAGTTGCAGAACCTAGAAAACCAACTATAC	780		
Db	721	GACCTTTAACAGTGGCTGAATTTGGAGCAGAGTTGCAGAACCTAGAAAACCAACTATAC	780		
QY	781	AAGAAATATTTCTTTACAGTGGAGAACCTACTTATCTGGGAAATGAAAACATCTGTTTTGG	840		
Db	781	AAGAAATATTTCTTTACAGTGGAGAACCTACTTATCTGGGAAATGAAAACATCTGTTTTGG	840		
QY	841	GCCAAACAGAAACAGAGCTCTGGTTTAAAGCCATATAAAAGATTTATTAACCCCTTCAACC	900		
Db	841	GCCAAACAGAAACAGAGCTCTGGTTTAAAGCCATATAAAAGATTTATTAACCCCTTCAACC	900		
QY	901	ACATTTGGCAACTAAAGATTTCTGTTGAGCTCTGGCAAAATTTTGATGAGTGAATGT	960		
Db	901	ACATTTGGCAACTAAAGATTTCTGTTGAGCTCTGGCAAAATTTTGATGAGTGAATGT	960		
QY	961	GCACAAACAGTTCATTTGTTTATTAATTTTGAATATGTGTTTTTACCTATGAAATATCCC	1020		
Db	961	GCACAAACAGTTCATTTGTTTATTAATTTTGAATATGTGTTTTTACCTATGAAATATCCC	1020		
QY	1021	TTTTTATTAATAATCATATGAGAAATCCCTTACCTATGAGAAACAAACACTCTCTGG	1080		
Db	1021	TTTTTATTAATAATCATATGAGAAATCCCTTACCTATGAGAAACAAACACTCTCTGG	1080		
QY	1081	TTTA	1084		
Db	1081	TTTA	1084		
RESULT 2	AF068227	4080 bp	mRNA	linear	PRI 28-JUL-1998
LOCUS	AF068227				
DEFINITION	Homo sapiens putative transmembrane protein (CLNS)		mRNA, complete cds.		
ACCESSION	AF068227				
VERSION	AF068227.1				GI:3342385
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
REFERENCE					
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L.				
JOURNAL	CLNS, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid				
MEDLINE	lipofuscinosis				
PUBMED	Nat. Genet. 19 (3), 286-288 (1998)				
REFERENCE	98324783				
AUTHORS	2 (bases 1 to 4080)				
TITLE	Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L.				
JOURNAL	Direct Submission				
FEATURES	Submitted (26-MAY-1998) Department of Human Molecular Genetics, National Public Health Institute, Mannerheimintie 166, Helsinki 00300, Finland				
source	Location/Qualifiers				
	1..4080				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/chromosome="13"				
	/map="13q22"				
	1..4080				
	/gene="CLNS"				
	1293..2516				
	/gene="CLNS"				
	/codon_start=1				
	/product="putative transmembrane protein"				
	/protein_id="AAC27614.1"				
	/db_xref="GI:3342386"				

/translation="MRNRRL1GPGSGADAGGAPBPBGALPAPMLLPPASOASGSGS
TGSIMAOBVDITDQAGMERMGAGAGASRSMCALALMLAVIPGMSRISJISRRHM
PVPKPRKPRKPDQOATFCPTSPSPVMEGDDDLVFLQAPVMEFKYGLDILG
HLKIMDAIPRSTLTGKNYTMWYELFQGNCTFPHLRBMQAPVPCMGACFEG
IDVHMKEKGTIVAVATISGNVOMAKMGTIYVETNMAFVMSPEKMEYED
SYOCSKVLRTFNKLAFFGAEPKNIETNRIFLYSEPTYLINETSVEFPTNKILG
LAIXRFPYPRKPLPTKEPFLSLILOIFDVAIVHKQPLFNFPEYVLPMPKPRIKLTY
EELPPLRNKTLISGL"

ORIGIN

Query Match 98.5%; Score 1067.8; DB 9; Length 4080;
Best Local Similarity 99.8%; Pred. No. 7,7e-235;
Matches 1069; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2Y 14 GAGCAGAGAGTGAACACAGGACAGGCGCCGAGATGCGCGCGGCGCGCGCTCGG 73
1443 GCGCAGAGAGTGAACACAGGACAGGCGCCGAGATGCGCGCGGCGCGCGCTCGG 1502
2Y 74 GAGCAGAGTGAACACAGGACAGGCGCCGAGATGCGCGCGGCGCGCTCGG 133
1503 GAGCAGAGTGAACACAGGACAGGCGCCGAGATGCGCGCGGCGCGCTCGG 1562
2Y 134 TCCCGGCTCTCGGCGATCCCTCCCGGCGCGCTCGGCGCGCTCGGCGCTCGG 153
1563 TCCCGGCTCTCGGCGATCCCTCCCGGCGCGCTCGGCGCGCTCGGCGCTCGG 1622
2Y 194 TTCGCTCGAAGACCTGATCTTATGTCAGCTAAGTATCTTCTGTCACCTGCTCA 253
1623 TTCGCTCGAAGACCTGATCTTATGTCAGCTAAGTATCTTCTGTCACCTGCTCA 1682
2Y 254 CCTATCCAGTATGAGGAGGATGATGACATGAACTTTTGGATTAACAAGCCCGCTA 313
1683 CCTATCCAGTATGAGGAGGATGATGACATGAACTTTTGGATTAACAAGCCCGCTA 1742
2Y 314 TGGGAATTTAAATGAGAGCTCTCGGACACCTTGAATTAATGATGATGATGATG 373
1743 TGGGAATTTAAATGAGAGCTCTCGGACACCTTGAATTAATGATGATGATGATG 1602
2Y 374 TTGCAAGTACATTAATCTGCGAAGACCTAACAATGATGATGATGATGATGATG 423
1803 TTGCAAGTACATTAATCTGCGAAGACCTAACAATGATGATGATGATGATGATG 1662
2Y 434 GGGCACTGATCTTCCGACCTCTCGAAGTGAATGATGATGATGATGATGATGATG 453
1863 GGGCACTGATCTTCCGACCTCTCGAAGTGAATGATGATGATGATGATGATGATG 1622
2Y 494 GGGCTGCTGCTTTTGAAGGAAATGATGATGATGATGATGATGATGATGATGATG 553
1923 GGGCTGCTGCTTTTGAAGGAAATGATGATGATGATGATGATGATGATGATGATG 1582
2Y 554 GTTCAAGTACATTAATCTGCGAAGACCTAACAATGATGATGATGATGATGATG 613
1983 GTTCAAGTACATTAATCTGCGAAGACCTAACAATGATGATGATGATGATGATG 2042
2Y 614 GAGCAATGAGAGAGATTTATGAGACATGAGATGATGATGATGATGATGATGATG 673
2043 GAGCAATGAGAGAGATTTATGAGACATGAGATGATGATGATGATGATGATGATG 2102
2Y 674 GCAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 723
2103 GCAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2162
2Y 734 TTGCTGATGATGAGAGTGAACAGATGAGAGATGAGAGATGAGAGATGAGAGAT 753
2163 TTGCTGATGATGAGAGTGAACAGATGAGAGATGAGAGATGAGAGATGAGAGAT 2222
2Y 794 TAGAGTGAAGAGAGTGAACAGATGAGAGATGAGAGATGAGAGATGAGAGATG 853
2223 TAGAGTGAAGAGAGTGAACAGATGAGAGATGAGAGATGAGAGATGAGAGATG 2282
2Y 854 AAGACTCTGCTTGAAGCAATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 913
2283 AAGACTCTGCTTGAAGCAATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 2342

QY 914 AAGAAATTTCTGTTGATGCTCTTGCAGAAATTTTGAAGCAAGTATGTCAGCAACAGTTG 973
2343 AAGAAATTTCTGTTGATGCTCTTGCAGAAATTTTGAAGCAAGTATGTCAGCAACAGTTG 2402
QY 974 TATTTGTTTATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1033
2403 TATTTGTTTATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2462
QY 1034 ACATATGAGAGAAATCCCTTACCTATCAGAGAAACAAACACTCTGTTTGA 1084
2463 ACATATGAGAGAAATCCCTTACCTATCAGAGAAACAAACACTCTGTTTGA 2513
Db 2463 ACATATGAGAGAAATCCCTTACCTATCAGAGAAACAAACACTCTGTTTGA 2513

RESULT 3
BD136338 1751 bp DNA linear PAT 18-SEP-2002
LOCUS 95 human secretory proteins.
DEFINITION BD136338
ACCESSION BD136338.1 GI:23231283
VERSION JP 2002506627-A/25.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM
Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1751)
Ruben,S.M., Ni,J., Rosen,C.A., Yu,G.L., Young,P.E., Fen,P.,
Soppet,D.R., Wei,Y.F., Endress,G.A., Duan,R.D., Kyaw,H., Ebner,R.,
Lafleur,D.W., Olsen,H.S., Shi,Y. and Moore,P.A.
95 human secretory proteins
Patent: JP 2002506627-A/25 05-MAR-2002.
TITLE
JOURNAL
HUMAN GENOME SCIENCES INC
COMMENT
OS Homo sapiens (human)
PN JP 2002506627-A/25
PD 05-MAR-2002
PE 18-MAR-1998 JP 200053673
PR 19-MAR-1998 US 60/078566,19-MAR-1998 US 60/078576 PR
19-MAR-1998 US 60/078573,19-MAR-1998 US 60/078574 PR
19-MAR-1998 US 60/078579,19-MAR-1998 US 60/078578 PR
19-MAR-1998 US 60/078581,19-MAR-1998 US 60/078577 PR
19-MAR-1998 US 60/078563,01-APR-1998 US 60/080314 PR
01-APR-1998 US 60/080312,01-APR-1998 US 60/080313 PI
M RUBEN, JIAN NI, CRAIG A ROSEN, GUO
LIANG YU, PAUL E YOUNG,
PI DANIEL R SOPPET, YING FEI WEI, GREGORY A ENDRESS, ROXANNE D DUAN,
PI HUA KYAW,
PI REINHARD EBNER, DAVID W LAFLAUR, HENRIK S OLSEN, YANGSU SHI, PAUL
PI A MOORE
PC C12N15/09, A61K38/00, A61K48/00, A61P43/00, C07K14/47, C07K16/18,
PC C12N1/15,
PC C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/68, G01N33/53, G01N33/56,
PC C12N15/00, A61K37/02, C12N5/00
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
FH Key Location/Qualifiers
FT source 1..1751
FT location/Qualifiers
FT /organism="Homo sapiens (human)".

FEATURES

source
1..1751
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 98.4%; Score 1067; DB 6; Length 1751;
Best Local Similarity 99.8%; Pred. No. 9.4e-235;


```

Db      379 TTCAGAGTACATTAAGTGGCAAGACTACACATGGAATGGATGACTTTCCAACTT 438
Qy      434 GGCAATGACATTTTCCCATCTCCGACCTGAAATGGATGCCCCCTTGGGTATCA 493
Db      439 GGCAATGACATTTTCCCATCTCCGACCTGAAATGGATGCCCCCTTGGGTATCA 498
Qy      494 GGCGCTGCTGCTTTTGGAGGAATTTGATGATGCTCATCGGAAGGAATGGACATTA 553
Db      499 GGCGCTGCTGCTTTTGGAGGAATTTGATGATGCTCATCGGAAGGAATGGACATTA 558
Qy      554 GTTCAAGTACCACTATATCAGAAACATGTTCAACCAATGGCAAGTGGGTGAACAG 613
Db      559 GTTCAAGTACCACTATATCAGAAACATGTTCAACCAATGGCAAGTGGGTGAACAG 618
Qy      614 GACAAATGAACAGGAATTTATTTATGACATGGAATGTAAGCCAGCCCAAGAAAGGGG 673
Db      619 GACAAATGAACAGGAATTTATTTATGACATGGAATGTAAGCCAGCCCAAGAAAGGGG 678
Qy      674 GCAGAGACATGTTGATTCCTACGACTGTCCTCAATTTGTGTTAAGACCTTTAACAAG 733
Db      679 GCAGAGACATGTTGATTCCTACGACTGTCCTCAATTTGTGTTAAGACCTTTAACAAG 738
Qy      734 TTGGCTGAATTTGACAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTT 793
Db      739 TTGGCTGAATTTGACAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTT 798
Qy      794 TACAGTGAAGAACCTTATCTGAGGAATGAAACATCTGTTTGGGCGCAACAGAAAC 853
Db      799 TACAGTGAAGAACCTTATCTGAGGAATGAAACATCTGTTTGGGCGCAACAGAAAC 858
Qy      854 AAGACTCTGTTGTTAGCCATAAAAGATTTTATCCCTCAACACCATTTGGCACT 913
Db      859 AAGACTCTGTTGTTAGCCATAAAAGATTTTATACCCCTTCAACACCATTTGGCACT 918
Qy      914 AAAGAAATTTCTGTAGTCTCTTGCATAATTTTGTATGACAGTATGTGCAACAAACAGTTC 973
Db      919 AAAGAAATTTCTGTAGTCTCTTGCATAATTTTGTATGACAGTATGTGCAACAAACAGTTC 978
Qy      974 TATTTGTTTATATTTTGAATTTGTTTATGCTTATGCAATGCAATTCCTTTATTTAAATA 1033
Db      979 TATTTGTTTATATTTTGAATTTGTTTATGCTTATGCAATGCAATTCCTTTATTTAAATA 1038
Qy      1034 ACATATGAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGCTTTA 1084
Db      1039 ACATATGAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGCTTTA 1089

RESULT 5
AK075109      2120 bp      mRNA      linear      PRI 03-SEP-2002
LOCUS      Homo sapiens cDNA FLJ90628 f1s, clone PLAC1003407, highly similar
DEFINITION      to Homo sapiens putative transmembrane protein (CLN5) mRNA.
ACCESSION      AK075109
VERSION      AK075109.1 GI:22760983
KEYWORDS      oligo capping; f1s (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS      1
              Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
              Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
              Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
              Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y.,
              Aotsuma,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
              Niimiya,K.
              NEDO human cDNA sequencing project
              Unpublished
              2 (bases 1 to 2120)
              Isogai,T. and Otsuki,T.
              Direct Submission
              Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
              Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

```

```

COMMENT
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
Institute of Medical Science, University of Tokyo, Laboratory of
Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
sequencing and clone selection: Helix Research Institute (supported
by Japan Key Technology Center etc.).

FEATURES
Source
1..2120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLAC1003407"
/issue_type="placenta"
/clone_lib="PLAC1"
/notes="Cloning vector: pME18FL3"

ORIGIN
Query Match      98.2%; Score 1064.6; DB 9; Length 2120;
Best Local Similarity 99.6%; Pred.No.3.5e-234;
Matches 1067; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

14 GGGCAGAGGTAGACACGCGCACAGCGCGCGAGATGCGCGCGCGCGCGCTCGG 73
19 GCGCAGAGGTAGACACGCGCACAGCGCGCGAGATGCGCGCGCGCGCGCTCGG 78
74 GAGCGCGCTTCTGCTGCTGCGCGCGCTGCGCTGCTTGGCTGCGCGCTGCG 133
79 GAGCGCGCTTCTGCTGCTGCGCGCGCTGCGCTGCTTGGCTGCGCGCTGCG 138
124 TCCGGGGTCTGGGGATCCCTCCCGGGCGCACTGCGCGGGTCCCTACAAAGGCTTGA 193
139 TCCGGGGTCTGGGGATCCCTCCCGGGCGCACTGCGCGGGTCCCTACAAAGGCTTGA 198
194 TTCCGTCACAAAACCTGATCCTTATTTGCAAGTAAATATCTTCTGTCACACTGCTCA 253
199 TTCCGTCACAAAACCTGATCCTTATTTGCAAGTAAATATCTTCTGTCACACTGCTCA 258
254 CCTATCCCAATTATGAGAGGTGATGATGACATTTGAAGTTTTCGATTACAGCCCACTA 313
259 CCTATCCCAATTATGAGAGGTGATGATGACATTTGAAGTTTTCGATTACAGCCCACTA 318
314 TGGGAATTTAAATATGAGAGCTCCTCGGAGCACTGAAATATGATGATGATGATGATG 373
319 TGGGAATTTAAATATGAGAGCTCCTCGGAGCACTGAAATATGATGATGATGATGATG 378
374 TTCAGAGTACATTAAGTGGCAAGACTACACATGGAATGGATGACTTTCCAACTT 433
379 TTCAGAGTACATTAAGTGGCAAGACTACACATGGAATGGATGACTTTCCAACTT 438
434 GGCACCTGTACATTTTCCCATCTCCGACCTGAAATGATGCCCCCTTGGGTATCA 493
439 GGCACCTGTACATTTTCCCATCTCCGACCTGAAATGATGCCCCCTTGGGTATCA 498
494 GCGCGTGCCTGCTTTTGGAGGAATTTGATGATGCTCATCGGAAGGAATGGACATTA 553
499 GCGCGTGCCTGCTTTTGGAGGAATTTGATGATGCTCATCGGAAGGAATGGACATTA 558
554 GTTCAAGTACCACTATATCAGAAACATGTTCAACCAATGGCAAGTGGGTGAACAG 613
559 GTTCAAGTACCACTATATCAGAAACATGTTCAACCAATGGCAAGTGGGTGAACAG 618
614 GACAAATGAACAGGAATTTATTTATGACATGGAATGTAAGCCAGCCCAAGAAAGGGG 673
619 GACAAATGAACAGGAATTTATTTATGACATGGAATGTAAGCCAGCCCAAGAAAGGGG 678
674 GCAGAGACATGTTGATTCCTACGACTGTCCTCAATTTGTGTTAAGACCTTTAACAAG 733
679 GCAGAGACATGTTGATTCCTACGACTGTCCTCAATTTGTGTTAAGACCTTTAACAAG 738
734 TTGGCTGAATTTGACAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTT 793

```

Db 739 TTGGCTGAATTTGGAGCAGATTCAAGAACATAGAAACCAACTATACAAATATTTCTT 798
QY 794 TACAGTGGAGAACTACTATATCGGGAAATGAAACATCTGTTTGGGCGAACAGAAAC 853
Db 799 TACAGTGGAGAACTACTATATCGGGAAATGAAACATCTGTTTGGGCGAACAGAAAC 858
QY 854 AAGACTCTGTTTGGAGCCATAAAAGATTATTAACCCCTTCAAAACAATTTGGCAACT 913
Db 859 AAGACTCTGTTTGGAGCCATAAAAGATTATTAACCCCTTCAAAACAATTTGGCAACT 918
QY 914 AAGAAATTTCTGTGAGTCTCTTGCAGAAATTTTGTATGATGATGTCAGCAACAAGTTC 973
Db 919 AAGAAATTTCTGTGAGTCTCTTGCAGAAATTTTGTATGATGATGTCAGCAACAAGTTC 978
QY 974 TATTTGTTTATTAATTTGATATGTTTGTATTTTACCTAGTAATTTCCCTTTATTAATA 1033
Db 979 TATTTGTTTATTAATTTGATATGTTTGTATTTTACCTAGTAATTTCCCTTTATTAATA 1038
QY 1034 ACATATGAAGAAATCCCTTTACCTATCAAGAAACAAACACTCTCTGTGTTTA 1084
Db 1039 ACATATGAAGAAATCCCTTTACCTATCAAGAAACAAACACTCTCTGTGTTTA 1089

RESULT 6

BD073413 1038 bp DNA linear PAT 27-AUG-2002
LOCUS Secreted protein which human chromosome 13 encodes.
DEFINITION BD073413

BD073413.1 GI:22619016
KEYWORDS JP 2001511345-A/12.
SOURCE JP 2001511345-A/12.
ORGANISM JP 2001511345-A/12.
artificial construct
artificial sequences.

REFERENCE 1 (bases 1 to 1038)
AUTHORS Sheppard, P.O. and Gilberton, D.G.
TITLE Secreted protein which human chromosome 13 encodes
JOURNAL Patent: JP 2001511345-A 12 14-AUG-2001;
ZYMOGENETICS INC

COMMENT

OS Artificial Sequence
PN JP 2001511345-A/12

PD 14-AUG-2001
PE 24-JUL-1998 JP 2000504249
PR 24-JUL-1997 US 60/053613

PI PAUL, O SHEPPARD, DIBRA G GILBERTON
PC C12N15/09, A61K38/00, A61K48/00, C07K14/47, C07K16/18, C12N1/15, PC
C12N1/19,

PC C12N1/21, C12N5/10, C12P21/02, C12Q1/68, C12N15/00, A61K37/02, C12N5/ PC
00

CC Degenerate nucleotide sequence encoding zsig46 polypeptide of

CC SEQ ID NO:2

FT source 1.1038 location/Qualifiers

FT location/Qualifiers
1.1038
/organism="Artificial Sequence".
/mol_type="genomic DNA"
/db_xref="taxon:32630"

FEATURES
source
1.1038
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN

Query Match 71.3%; Score 772.4; DB 6; Length 1038;
Best Local Similarity 61.7%; Pred. No. 5, 2e-167;
Matches 640; Conservative 232; Mismatches 165; Indels 0; Gaps 0;

QY 47 ATGCGGCGGGCGCGGGCGGCTGGGAGCGCGCTTCTGCTGGCGCTGGCGCTG 106
Db 1 ATGCGGCGGGCGCGGGCGGCTGGGAGCGCGCTTCTGCTGGCGCTGGCGCTG 60
QY 107 CTTTGGCGCGGGCGGCTGGCGGGCGGCTGGCGGGCGGCTGGCGGGCGGCTG 166
Db 61 YNTGTGTTGCGTGTGTTGCGTGTGTTGCGTGTGTTGCGTGTGTTGCGTGTGTT 120

QY 167 TGGCGGCGGGCGGCTGGGAGCGCGCTTCTGCTGGCGCTGGCGCTGGCGCTG 226
Db 121 TGGCGGCGGGCGGCTGGGAGCGCGCTTCTGCTGGCGCTGGCGCTGGCGCTG 180
QY 227 AAGTATCTTCTGTCCAACTGGGCTCACTTATCCCACTTATGAGGCGGATGATGACAT 286
Db 181 AAGTATCTTCTGTCCAACTGGGCTCACTTATCCCACTTATGAGGCGGATGATGACAT 240
QY 287 GAACTTTTGCATTAACAAGCCCGAGTATGGAATTTTAATATGAGAGCTCTGGGACAC 346
Db 241 GAACTTTTGCATTAACAAGCCCGAGTATGGAATTTTAATATGAGAGCTCTGGGACAC 300
QY 347 TTGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406
Db 301 TTGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 407 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 466
Db 361 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 467 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 526
Db 421 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 527 GTTCACTGGAAGAAATGAGCACTTATGATGATGATGATGATGATGATGATGATGATGAT 586
Db 481 GTTCACTGGAAGAAATGAGCACTTATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 587 AACCAATGGAAGAAATGAGCACTTATGATGATGATGATGATGATGATGATGATGATGAT 646
Db 541 AACCAATGGAAGAAATGAGCACTTATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 647 AATGTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 706
Db 601 AATGTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 660
QY 707 AATGTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 766
Db 661 AATGTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 720
QY 767 GAAACCAATGGAAGAAATGAGCACTTATGATGATGATGATGATGATGATGATGATGAT 826
Db 721 GAAACCAATGGAAGAAATGAGCACTTATGATGATGATGATGATGATGATGATGATGAT 780
QY 827 ACATCTGTTTGGGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 886
Db 781 ACATCTGTTTGGGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 840
QY 887 TACCCCTTCAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 946
Db 841 TACCCCTTCAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 900
QY 947 GATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1006
Db 901 GATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 1007 CCTTGAATTTCTGTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 1066
Db 961 CCTTGAATTTCTGTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 1020

QY 1067 AAAACACTCTGCTGTTT 1083
Db 1021 AAAACACTCTGCTGTTT 1037

RESULT 7

BD025487

LOCUS BC025487 2318 bp mRNA linear ROD 16-APR-2003
DEFINITION Mus musculus ceroid-lipofuscinosis, neuronal 5, mRNA (cDNA clone
IMAGE:5251891), partial cds.

ACCESSION BC025487

VERSION BC025487.1 GI:19343790

KEYWORDS


```

Qy 811 TTATCTGGGAAATGAAACATCTGTTTTTGGGCCAAGAGAAACAGACTTGGTTAGC 870
Db 737 TTACTGGGAAATGAAACATCTATTTTGGGCCCAAGAGAAACAGACTTGGTTAGC 796
Qy 871 CATAAAGATTATTAATCCCTTCAACCAATTCGCCAATAAGATTTCTGTTAG 930
Db 797 CATAAAGATTATTAATGGCCCTTCAAGCCGATTTTCAACCAAGATTTCTGATGA 856
Qy 931 TCTCTGGAATTTTGTATGAGATGTGTGACAAAGATTTCTGTTTAAATTT 990
Db 857 TTTCTGGAATTTTGTATGAGATGTGTGACAAAGATTTCTGTTTAAATTT 916
Qy 991 TGAATATTTGTTTATTTACCTATGAATTCCTTTTATTAATAATACATGAAGAAATCCC 1050
Db 917 TGAGATTTGTTTATTTACCTATGAATTCCTTTTATTAATAATACATGAAGAAATCCC 976
Qy 1051 TTTACCTATGAGAAACAAACACTCTCTGTTT 1083
Db 977 GTTACCTATGAGAAACAAACACTCTCTGTTT 1009

RESULT 8
AC109554/c 198524 bp DNA linear HTG 11-OCT-2002
LOCUS Rattus norvegicus clone CH230-331D3, *** SEQUENCING IN PROGRESS
DEFINITION ***
AC109554
AC109554 GI:23820683
VERSION HTG_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 198524)
Muzny,D.,Marle,M.,Metzker,M.,Lee,A.,Abramson,S.,Adams,C.,Alder,J.,
Allen,C.,Allen,H.,Albrooks,S.,Amin,A.,Anguiano,D.,
Anyalebechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,
Baldwin,D.,Bandaranaike,D.,Barber,M.,Barnstead,M.,Benahmed,F.,
Biswal,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,
Bryant,N.,Bunay,C.,Burch,P.,Burrell,K.,Calderson,E.,
Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,
Chacko,J.,Chavez,D.,Chen,R.,Chen,Y.,Chen,Z.,Chu,C.,
Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,
Davila,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dederich,D.,
Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Divya,K.,
Diaper,H.,Dugan-Rocha,S.,Dunn,A.,Dutkin,K.,Duvall,B.,Eaves,K.,
Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,
Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Foster,P.,
Frazer,C.M.,Gabisi,A.,Ganta,R.,Garcia,A.,Garner,T.,Gatta,M.,
Gebregeorgis,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,
Gunaratne,P.,Haaland,W.,Hami,C.,Hamilton,C.,Hamilton,K.,
Harvey,I.,Havlik,P.,Hawes,A.,Henderson,N.,Hernandez,J.,
Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,M.,
Hollins,B.,Howells,S.,Huiy,K.S.,Hume,J.,Idlebird,D.,Jackson,A.,
Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jolyvet,A.,
Karpathy,S.,Kelly,S.,Kelly,S.,Khan,Z.,King,L.,Koyar,C.,
Kowis,C.,Kraft,C.L.,Lebow,H.,Levan,J.,Lewis,L.,Li,Z.,Liu,J.,
Liu,J.,Liu,W.,Liu,Y.,London,P.,Longacre,S.,Lopez,J.,
Lorenshew,L.,Louisege,D.,Lozano,R.J.,Lu,X.,Ma,J.,
Maheshwari,M.,Mahindaratne,M.,Maimoud,M.,Malloy,K.,Mangum,A.,
Mangum,B.,Mapua,P.,Martin,K.,Martin,R.,Martinez,E.,
Mawhinney,S.,McLeod,M.P.,McNeill,T.Z.,Meenen,E.,
Millsavljetic,A.,Miner,G.,Munja,E.,Montemayor,J.,Moore,S.,
Morjan,M.,Morris,K.,Morris,S.,Mundasa,M.,Murphy,M.,Nair,L.,
Nankervyls,C.,Neal,D.,Newton,G.,Olanpinuysoon,A.,Pal,S.,Parks,K.,
Pasceriak,S.,Paul,H.,Perez,A.,Perez,L.,Pfannkuch,C.,
Plopper,F.,Polindexter,A.,Popovic,D.,Prims,B.,Pu,L.-L.,
Puzo,M.,Quiroz,J.,Rachlin,E.,Reeves,K.,Regier,M.A.,Reish,R.,
Reilly,B.,Reilly,M.,Ren,Y.,Reuter,M.,Richards,S.,Riggs,F.,
Rivers,C.,Rodkey,T.,Rojas,A.,Rose,M.,Rocha,R.,Ritz,S.U.,
Sanders,W.,Savery,G.,Scherer,S.,Scott,G.,Shatsman,S.,Shen,H.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smaiz,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Swack,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Treijs,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,D., Walker,B., Wang,U.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Wilson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhan,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 198524)
Worley,K.C.
Direct Submission
Submitted (05-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 198524)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 11, 2002 this sequence version replaced gi:21738217.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOLY
Center clone name: CH230-331D3
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 177481 bases at least Q40
Consensus quality: 179640 bases at least Q30
Consensus quality: 181141 bases at least Q20
Estimated insert size: 183954; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOT: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 198524: contig of 198524 bp in length.
Location/Qualifiers
1. 198524
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-331D3"
1. 1183
misc_feature

```

```

misc_feature
131409..131443
/note="clone boundary
clone end:17
site:Md01
end_sequence:RXAPC14TV"
197370..198217
/note="clone boundary
clone end:5p6
site:Md01
end_sequence:RXAPC14TV"

ORIGIN

Query Match      61.1%; Score 662.8; DB 2; Length 198524;
Best Local Similarity 77.0%; Pred. No. 3.4e-141;
Matches 808; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

2y 30 CGGCAAGAGGCGCGGAGATGCGGCGGCGGCGGCGGCTCGGAGACGGCTTCTGTGT 89
Db 38682 CGAGCGCCCTGGCCCGGGAGCCACCGGAGCTTCAGGGAGTCCGTGGGCGCAGGTGGAGCGC 38623
2y 90 GCTGGGCGCTGGCGCTGCTTGGCTGGCGGCTGGCTTCCGGGCTTCCTCCGGCTTCGGGCA 149
Db 38622 ACTGGCGCGAGGCGCTCGCGCTGGCGCTACTGGGCGCTGGCGGCGCACTCGGGCGCGTCCC 38563
2y 150 TCCCGTCCCGCGCGGCGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 209
Db 38562 CGAGGCTGGGCGAGCGCTGGCGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 38503
2y 210 ATCTTATTTGCTCAAGTAACTATCTTCTGTCCTCAATGCTGCTCACTATCCAGTTATGG 269
Db 38502 ATCCCTTCTGTCAAGCCAGATATCTTCTGCTTCTGCTGCTCACTATCCAGTTATGG 38443
2y 270 AGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 329
Db 38442 AGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 38383
2y 330 GAGACCTCCCGGCGGCACTTGTAAATTTATGATGATGATGATGATGATGATGATGATGATG 369
Db 38382 GAGACCTCCCGGCGGCACTTGTAAATTTATGATGATGATGATGATGATGATGATGATGATG 38323
2y 390 CTGGCAAGAACTACACATGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 449
Db 38322 CCGGCAAGAACTACACATGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 38263
2y 450 CCGATCTCGGAGCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 509
Db 38262 CCGATCTCGGAGCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 38203
2y 510 TTGAGGGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 569
Db 38202 TTGAGGGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 38143
2y 570 TATCAGGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 629
Db 38142 TATCAGGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 38083
2y 630 TTTATTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 689
Db 38082 TTTATTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 38023
2y 690 ATTCTTCAAGCTGTTCCAAATTTGTTGTTAGAGCTTTTAAAGAGTTGGCTGAAATTTGGAG 749
Db 38022 ATTCTTCAAGCTGTTCCAAATTTGTTGTTAGAGCTTTTAAAGAGTTGGCTGAAATTTGGAG 37963
2y 750 CAGAGTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 809
Db 37962 CAGAGTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 37903
2y 810 CTTATCTGGGAAATGAAACCTGTTTGGGCGCAAGGAAACCAAGACTTTGGTTTGG 869
Db 37902 TTTACTTGGGAAATGAAACCTGTTTGGGCGCAAGGAAACCAAGACTTTGGTTTGG 37843

```

```

Qy 870 CCATTAATAAGATTTTATTTATCCCTTCAACACATTTGCGCACTAATGATTTCTGTGA 929
Db 37842 CCATTAATAATAATTTTATGCGGCCGCTCAACGGAATTTGTCAACCAAGATTTCTGTGA 37783
Qy 930 GTCTCTGCAAAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 989
Db 37782 ATTTCTGCAAAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 37723
Qy 990 TTGAATATGATTTTATCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1049
Db 37722 TTGAATATGATTTTATCTATGATGATGATGATGATGATGATGATGATGATGATGATG 37663
Qy 1050 CTTTACTTATCCAGAACCAAAACCTCTCTG 1079
Db 37662 CTTTACTTATCCAGAACCAATACGCAATTTTACCG 37633

RESULT 9
AC107510
LOCUS
DEFINITION
Rattus norvegicus clone CH230-138J6, *** SEQUENCING IN PROGRESS
***, 4 unordered pieces.
ACCESSION
AC107510
VERSION
GI:30580095
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine;
Rattus.

1 (bases 1 to 227920)
Nuzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cessat, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, P.,
Fraser, C. M., Gabisi, A., Gant, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, X., Gill, R., Grady, M., Guerra, T., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Huily, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Lervan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshuwa, L., Louieged, H., Lozano, R. J., Lu, X., Ma, T.,
Maheshwari, M., Mahindaratne, M., Mahmood, M., Mallory, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munkata, M., Murphy, M., Nair, L.,
Nankervill, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwakoelimeh, O., Okunolu, G., Olarinmuyegoon, A., Pal, S., Parks, K.,
Pasernak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C.,
Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quirio, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Silter, C. D., Smaj, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,

```

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wleczek, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, O., Yoon, L., Yoon, V., Yu, F., Zhang, U., Zhou, J., Zhou, X., Zhou, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 227920)

Worley, K.C.
Direct Submission
Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 227920)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:22855855. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GJBR
Center clone name: CH230-138J6
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 214528 bases at least Q40
Consensus quality: 219064 bases at least Q20
Estimated insert size: 223690; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafi_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 223559: contig of 223559 bp in length
* 223560 223659: gap of unknown length
* 223660 224738: contig of 1079 bp in length
* 224739 224838: gap of unknown length
* 224839 226289: contig of 1451 bp in length
* 226290 226389: gap of unknown length
* 226390 227920: contig of 1531 bp in length.
Location/Qualifiers
1..227920
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"

misc_feature /clone="CH230-138J6"
1..1145
misc_feature /note="wgs contig"
222227..223559
misc_feature /note="wgs contig"

ORIGIN
Query Match 58.7%; Score 635.8; DB 2; Length 227920;
Best Local Similarity 77.0%; Pred. No. 5,8e+135;
Matches 775; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

30 CGGACAGAGGCGCCGAGATGCGGCGGCGCGCGCGCTTCGGAGCGCTTCGTGT 89
37202 CGAGCCCTCGGCGCGGAGCCAGCGGACTTCAGGAGTCCGTGGCCAGGTGAGCGC 37261
90 GGTGGGCGCTGCGCTGCTTGGCTGCGGCGGCTTCGGGCTTCGGGCTTCGGGCA 149
37262 ACTGGCGCGAGCGCTCGCGCTGCGGCTTCAGGCGGCGGCGGCGGCGGCTCC 37321
150 TCCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 209
37322 CGAGCTTCGGGCGGAGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 37381
210 ATCTTATTCAGCTAGTATGATCTTCTGCTCACTGCTCACTGCTCACTGCTTATG 269
37382 ATCCCTCTGTCAAGCAAGTATGATCTTCTGCTCACTGCTCACTGCTCACTG 37441
270 AGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 329
37442 AGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 37501
330 GAGACCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 389
37502 GAGACCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 37561
390 CTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 449
37562 CGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 37621
450 CCATCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 509
37622 CCATCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 37681
510 TTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 569
37682 TTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 37741
570 TATCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 629
37742 TATCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 37801
630 TTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689
37802 TTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 37861
690 ATTCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 749
37862 ATTCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 37921
750 CAGAGTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 809
37922 CAGAGTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 37981
810 CTATATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 869
37982 TTATATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 38041
870 CCAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 929
38042 CCAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 38101
930 GTCTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 989

Db 38102 ATTCTTGAAATTTTTCACACAGTATATGACAGAGAGTTCTACCTGTTTATTAACCT 38161
QY 990 TTGAATATTTGGTTTTTACCTATGAAATTCCTTTTATTAATAACA 1036
Db 38162 TTGAGATTGGTTTCTACCTATGAAACCCCTTTTCTGCTAGGA 38208

RESULT 10
BD125187 697 bp DNA linear PAT 18-SEP-2002
LOCUS BD125187
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD125187
VERSION BD125187.1 GI:23220132
KEYWORDS JP 2002017375-A/618.
SOURCE JP 2002017375-A/618.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 697)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 618 22-JAN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/618
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUO OTSUKI, HISASHI KOGA
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FT source 1..697
Location/Qualifiers
/organism="Homo sapiens (human)".
1..697
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
ORIGIN
Query Match 52.3%; Score 567; DB 6; Length 697;
Best Local Similarity 95.2%; Pred. No. 8,9e-120;
Matches 637; Conservative 0; Mismatches 24; Indels 8; Gaps 5;

QY 14 GGGCAGAGGATGACACGCGACAGCGCGCGAGATGCGCGCGCGCGCGCGCTCGG 73
Db 19 GCGCAGAGGATGACACGCGACAGCGCGCGCGAGATGCGCGCGCGCGCGCGCTCGG 78
QY 74 GGAGCGCGCTTCTGCTGCTGGCGCTGGCGCTGCTTGGCTCGCGGTGTTCCGGCTCGG 133
Db 79 GGAGCGCGCTTCTGCTGCTGGCGCTGGCGCTGCTTGGCTCGCGGTGTTCCGGCTCGG 138
QY 134 TCCGCGGCTCTGGCGCATCCCTCCCGCGCGCATGCGCGGCTTGAACAAGCGCTTGAAC 193
Db 139 TCCGCGGCTCTGGCGCATCCCTCCCGCGCGCATGCGCGGCTTGAACAAGCGCTTGAAC 198
QY 194 TTCCGCTCCAAAACCTGATCTTATTTGTCAGCTAAGTAACTTTTGTGTCAGCTGCTCA 253
Db 199 TTCCGCTCCAAAACCTGATCTTATTTGTCAGCTAAGTAACTTTTGTGTCAGCTGCTCA 258
QY 254 CCTATCCAGTTATGAGGGGTGATGATGACATGGAAGTTTTCATTAACAAGCCCAAGTA 3..3
Db 259 CCTATCCAGTTATGAGGGGTGATGATGACATGGAAGTTTTCATTAACAAGCCCAAGTA 3..8

QY 314 TGGGAATTTAAATATGAGACCTCTCGGACACTTGAATAATATGATGATGATGGA 373
Db 319 TGGGAATTTAAATATGAGACCTCTCGGACACTTGAATAATATGATGATGATGGA 378
QY 374 TTGGAAGTACATTAACTGGGACAGACTACCAATGGAATGATGAACTTTTCAACTT 433
Db 379 TTGGAAGTACATTAACTGGGACAGACTACCAATGGAATGATGAACTTTTCAACTT 438
QY 434 GCGAAGTGTACATTTTCCCATCTCCGACCTGGAATGGAATGCCCTTTCTGATGATCA 493
Db 439 GCGAAGTGTACATTTTCCCATCTCCGACCTGGAATGGAATGCCCTTTCTGATGATCA 498
QY 494 GCGCCTCGCTCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGAT 552
Db 499 GCGCCTCGCTCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGAT 558
QY 553 AGTTCAAGTACCAATATATC-AGGAAACATGTTCAACCAATATGCG-AAAATGGGTGAA 610
Db 559 AGTTCAAGTACCAATATATCAAGAAACATGTTCAACCAATATGCGCAAAAGTGGTGA 618
QY 611 CAGG--ACATGAAACAGGAATTTATTTGAGA--CATGGAATGTAAGACCGCCGAG 665
Db 619 CAAAGACATGGAACAAAGAAATTTATGAGAAATGGAATTTTAAAGCCACCCAA 678
QY 666 AAAAGGGG 674
Db 679 NAAAAANG 687

RESULT 11
BD126453 697 bp DNA linear PAT 18-SEP-2002
LOCUS BD126453
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD126453
VERSION BD126453.1 GI:23221398
KEYWORDS JP 2002017375-A/1884.
SOURCE JP 2002017375-A/1884.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 697)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 1884 22-JAN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/1884
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUO OTSUKI, HISASHI KOGA
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FT source 1..697
Location/Qualifiers
/organism="Homo sapiens (human)".
1..697
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
ORIGIN
Query Match 52.3%; Score 567; DB 6; Length 697;
Best Local Similarity 95.2%; Pred. No. 8,9e-120;

Matches	637	Conservative	0	Mismatches	24	Indels	8	Gaps	5
QY	14	GGGAGAGGAGGAGACAGGACAGGGGCGGAGATGGGCGGGCGGGCGGGCTCGG	73						
Db	19	GCGGAGAGGATGACAGGACAGGGGCGGAGATGGGCGGGCGGGCGGGCTCGG	78						
QY	74	GGAAGCGCTTCTGAGTCTGAGGCTGGGCGCTGGCTGGCTGGCTGGCTGGCTGG	133						
Db	79	GGAAGCGCTTCTGAGTCTGAGGCTGGGCGCTGGCTGGCTGGCTGGCTGGCTGG	138						
QY	134	TCCCGGGCTTCGGGACATCCCTCCCGGCGCACTGGCGGGTCCCTCAAGGCGCTT	193						
Db	139	TCCCGGGCTTCGGGACATCCCTCCCGGCGCACTGGCGGGTCCCTCAAGGCGCTT	198						
QY	194	TTCCGCTCAAAACCTGATCTCTTATGTCAGAGTAAAGTACTCTTCTGCTCAATG	253						
Db	199	TTCCGCTCAAAACCTGATCTCTTATGTCAGAGTAAAGTACTCTTCTGCTCAATG	258						
QY	254	CCTATCCCACTTATGAGGAGTGAATGATGATGATGATGATGATGATGATGATG	313						
Db	259	CCTATCCCACTTATGAGGAGTGAATGATGATGATGATGATGATGATGATGATG	318						
QY	314	TGGGAAATTTAAATATGAGACCTCTCTGGGACCTTGAAATTTATGCAATGATG	373						
Db	319	TGGGAAATTTAAATATGAGACCTCTCTGGGACCTTGAAATTTATGCAATGATG	378						
QY	374	TTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	433						
Db	379	TTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	438						
QY	434	GGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	493						
Db	439	GGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	498						
QY	494	GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	552						
Db	499	GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	558						
QY	553	AGTTCAAGTGAAGCACTATATC-AGGAAACATGTTCAACCAAAATGGC-AAA	610						
Db	559	AGTTCAAGTGAAGCACTATATC-AGGAAACATGTTCAACCAAAATGGC-AAA	618						
QY	611	CAGG--ACAATGAACAGGAAATTTATATGAGA--CANGAAATGTAAGGCGG	665						
Db	619	CAGGAAATGTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	678						
QY	666	AAAAGGGGG 674							
Db	679	AAAAAANGG 687							

RESULT 12
AC001226 106988 bp DNA linear PRI 29-MAY-1997
LOCUS AC001226
DEFINITION Genomic sequence from Human 13, complete sequence.
AC001226
VERSION AC001226.1 GI:2133862
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 106988)
Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,
Fasman, K.H. and Lander, E.S.
TITLE Genomic sequence from Human 13
JOURNAL Unpublished
2 (bases 1 to 106988)
Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,
Fasman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,
Barna, N., Brown, K., Cooke, P., Daly, M.J., Forrest, C., Frapp, M.J.,
Gage, D., Geratger, K., Hagos, B., Jacotot, L., Lane, M., Mackenzie, J.,
Marquis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A.,

TITLE	JOURNAL	REFERENCE	AUTHORS
Direct Submission	Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	3 (bases 1 to 106988)	Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W., Fasman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P., Barna, N., Chang, A., Cooke, P., Daly, M.J., Forrest, C., Frapp, M.J., Gage, D., Geratger, K., Hagos, B., Jacotot, L., Lane, M., Mackenzie, J., Marquis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, D., Stilwell, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and Zody, M.
TITLE	JOURNAL	COMMENT	
Direct Submission	Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	On May 29, 1997 this sequence version replaced gi:1932720. The Staden databases, finishing information, and all chromatographic files used in the assembly of this clone are available from our anonymous ftp site.	
FEATURES	source		
1..106988	Organism="Homo sapiens"		
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
/chromosome="13"			
/clone="76N15"			
/clone_lib="unknown"			
complement(113..200)			
/rpt_family="THR1B"			
complement(202..611)			
/rpt_family="L1PA3"			
complement(674..694)			
/rpt_family="AT_rich"			
complement(835..981)			
/rpt_family="MIR2"			
1632..1920			
/rpt_family="AluB"			
1970..2271			
/rpt_family="AluSp"			
2779..2946			
/rpt_family="FAM"			
3552..3828			
/rpt_family="AluB"			
4361..4640			
/rpt_family="AluSg"			
5253..5385			
/rpt_family="MLTIF"			
5914..6106			
/rpt_family="MLTIF"			
complement(6341..6435)			
/rpt_family="MER34"			
complement(6838..7140)			
/rpt_family="AluSx"			
7394..7455			
/rpt_family="MER5A"			
7464..7627			
/rpt_family="MER5B"			
complement(7624..7792)			
/rpt_family="MER5B"			
complement(8363..8391)			
/rpt_family="AT_rich"			
8480..8639			
/rpt_family="MER5B"			
complement(9199..9286)			
/rpt_family="TAN"			
9340..9644			

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

Location/Qualifiers
1..106988
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="76N15"
/clone_lib="unknown"
complement(113..200)
/rpt_family="THR1B"
complement(202..611)
/rpt_family="L1PA3"
complement(674..694)
/rpt_family="AT_rich"
complement(835..981)
/rpt_family="MIR2"
1632..1920
/rpt_family="AluB"
1970..2271
/rpt_family="AluSp"
2779..2946
/rpt_family="FAM"
3552..3828
/rpt_family="AluB"
4361..4640
/rpt_family="AluSg"
5253..5385
/rpt_family="MLTIF"
5914..6106
/rpt_family="MLTIF"
complement(6341..6435)
/rpt_family="MER34"
complement(6838..7140)
/rpt_family="AluSx"
7394..7455
/rpt_family="MER5A"
7464..7627
/rpt_family="MER5B"
complement(7624..7792)
/rpt_family="MER5B"
complement(8363..8391)
/rpt_family="AT_rich"
8480..8639
/rpt_family="MER5B"
complement(9199..9286)
/rpt_family="TAN"
9340..9644

```
/rpt_family="AluJo"
repeat_region 9880..9946
/rpt_family="MIR2"
repeat_region 10208..10787
/rpt_family="LIMB3"
repeat_region 10984..11297
/rpt_family="AluSg"
repeat_region 11311..11470
/rpt_family="L1PA8"
repeat_region 11471..11495
/rpt_family="AT_rich"
repeat_region 12394..12440
/rpt_family="AT_rich"
repeat_region 13782..13921
/rpt_family="MIR"
repeat_region 14035..14316
/rpt_family="AluSx"
repeat_region 15018..15046
/rpt_family="AT_rich"
repeat_region 15080..15373
/rpt_family="Aluub"
repeat_region 15388..15881
/rpt_family="AluJo"
repeat_region 18038..18071
/rpt_family="CAT)n"
repeat_region 20389..20434
/rpt_family="FAM"
repeat_region 21338..21534
/rpt_family="AluSx"
repeat_region 23331..23353
/rpt_family="GC_rich"
repeat_region 23546..23578
/rpt_family="GC_rich"
repeat_region 24565..24707
/rpt_family="MIR"
repeat_region 26013..26047
/rpt_family="MIR2"
repeat_region 27717..28532
/rpt_family="L1PB1"
repeat_region 28553..28854
/rpt_family="AluSg"
repeat_region 28860..29076
/rpt_family="L1"
repeat_region 29365..29561
/rpt_family="MIR"
repeat_region 30050..30147
/rpt_family="MIR"
repeat_region 31487..31812
/rpt_family="AluJo"
repeat_region 33629..33732
/rpt_family="MIR"
repeat_region 34434..34729
/rpt_family="AluSx"
repeat_region 35070..35361
/rpt_family="AluSx"
repeat_region 36437..36461
/rpt_family="AT_rich"
repeat_region 39967..40261
/rpt_family="AluSx"
repeat_region 40741..40915
/rpt_family="AluSg"
repeat_region 40963..41663
/rpt_family="L1PA2"
repeat_region 42042..42352
/rpt_family="MLT1E"
repeat_region 42467..42771
/rpt_family="AluSg"
repeat_region 42772..42848
/rpt_family="LIMB8"
repeat_region 42849..43151
/rpt_family="Aluub"
repeat_region 43152..43318
/rpt_family="L1MA10"
```

```
repeat_region 43253..43427
/rpt_family="LIMB8"
repeat_region 43429..43672
/rpt_family="LIMB8"
repeat_region 44188..44481
/rpt_family="AluSx"
repeat_region 44541..45246
/rpt_family="L1TR8"
repeat_region 46999..47032
/rpt_family="AT_rich"
repeat_region 47411..47431
/rpt_family="AT_rich"
repeat_region 47612..47684
/rpt_family="CAT)n"
repeat_region 47946..48239
/rpt_family="AluJo"
repeat_region 49827..49872
/rpt_family="AT_rich"
repeat_region 50160..50186
/rpt_family="AT_rich"
repeat_region 50952..51121
/rpt_family="MIR5A"
repeat_region 52140..52284
/rpt_family="AluJo"
repeat_region 53772..53974
/rpt_family="MIR1B"
repeat_region 54225..54346
/rpt_family="MIR"
repeat_region 54533..54579
/rpt_family="TGAA)n"
repeat_region 54580..54736
/rpt_family="Aluub"
repeat_region 54755..54782
/rpt_family="CA)n"
```

```
Query Match 47.3%; Score 512.2; DB 9; Length 106988;
Best Local Similarity 99.4%; Pred. No. 1.3e-106;
Matches 514; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 568 TATATCAGAAACATGTTCAACCAATGGCAAAAGGGTGAAACAGACATGAAACAG 627
Db 31859 TAAACTAGAAACATGTTCAACCAATGGCAAAAGGGTGAAACAGACATGAAACAG 31918

QY 628 AATTTATATGACATGCAATGTAAAGCCAGCCAGAAAGGGGGCAGACATGTT 687
Db 31919 AATTTATATGACATGCAATGTAAAGCCAGCCAGAAAGGGGGCAGACATGTT 31978

QY 688 TGAATCTAGACATGTTCCAAATTTGTAAAGACCTTTAACAAGTGGCTGAATTGG 747
Db 31979 TGAATCTAGACATGTTCCAAATTTGTAAAGACCTTTAACAAGTGGCTGAATTGG 32038

QY 748 AGCAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTTACATGGAGAAC 807
Db 32039 AGCAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTTACATGGAGAAC 32098

QY 808 TACTATATCGGAAAGAAACATCTGTTTTGGGCCCAACAGGAAACAAGCTCTGGTT 867
Db 32099 TACTATATCGGAAAGAAACATCTGTTTTGGGCCCAACAGGAAACAAGCTCTGGTT 32158

QY 868 AGCATAAAGAGTTTATATTAACCCCTCAACACAGATTGGCCAACTAAGAAATTCGTT 927
Db 32159 AGCATAAAGAGTTTATATTAACCCCTCAACACAGATTGGCCAACTAAGAAATTCGTT 32218

QY 928 GAGTCTCTGCAAAATTTTGATGACAGTATGTGCAAAACAGTCTATTTGTTTATAA 987
Db 32219 GAGTCTCTGCAAAATTTTGATGACAGTATGTGCAAAACAGTCTATTTGTTTATAA 32278

QY 988 TTTTGAATATGTTTATTAACCTATGAAATTCCTTTTATTAATAACATAGAAAT 1047
Db 32279 TTTTGAATATGTTTATTAACCTATGAAATTCCTTTTATTAATAACATAGAAAT 32338

QY 1048 CCTTACCTATCAGAAACAAACACTCTGTTTA 1084
```

Db 32339 CCCTTACCTATCAGAAACAAACACTCTCTGTTTA 32375

RESULT 13
AL136440 169362 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.
DEFINITION AL136440
ACCESSION AL136440
KEYWORDS HTG, HTGS_PHASE1, HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Sep 8, 2000 this sequence version replaced gi:6982057.

COMMENT
----- Genome Center
Center: Sanger Centre
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA185124
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 16208 bases at least Q40
Consensus quality: 164521 bases at least Q30
Consensus quality: 16040 bases at least Q20
Insert size: 168062; sum-of-contigs
Insert size: 164936; 1.7% error; agarose-fp
Quality coverage: 4.08x in Q20 bases; sum-of-contigs Quality
coverage: 4.16x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3400: contig of 3400 bp in length
* 3500: gap of 100 bp
* 16444: contig of 12944 bp in length
* 16445: gap of 100 bp
* 16445: contig of 2970 bp in length
* 19515: gap of 100 bp
* 19515: contig of 2627 bp in length
* 22241: gap of 100 bp
* 22242: contig of 3185 bp in length
* 25526: gap of 100 bp
* 25527: contig of 7444 bp in length
* 25627: gap of 100 bp
* 33070: contig of 7444 bp in length
* 33071: gap of 100 bp
* 33171: contig of 34450 bp in length
* 67720: gap of 100 bp
* 67721: contig of 25693 bp in length
* 67721: gap of 100 bp
* 93513: gap of 100 bp
* 93514: contig of 20568 bp in length
* 114081: gap of 100 bp
* 114082: contig of 12025 bp in length
* 114182: gap of 100 bp
* 126306: gap of 100 bp
* 126307: contig of 18277 bp in length
* 144583: gap of 100 bp
* 144584: contig of 2453 bp in length
* 147136: gap of 100 bp
* 147137: contig of 17689 bp in length
* 147236: gap of 100 bp
* 164925: contig of 17689 bp in length

* 164926 165025: gap of 100 bp
* 165026 169362: contig of 4337 bp in length.

FEATURES
Source
1. 169362
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-185124"
/clone_1b="RP11-11.1"
1. 3400
/note="assembly_fragment:00318
clone end:77
vector side:left"

misc_feature
3501..16444
/note="assembly_fragment:00726
fragment_chain:1"

misc_feature
16545..19514
/note="assembly_fragment:00157
fragment_chain:1"

misc_feature
19615..22241
/note="assembly_fragment:00846
fragment_chain:1"

misc_feature
22342..25526
/note="assembly_fragment:00271
fragment_chain:1"

misc_feature
25627..33070
/note="assembly_fragment:01555
fragment_chain:1"

misc_feature
33171..67620
/note="assembly_fragment:00768"
67721..93413
/note="assembly_fragment:00828"

misc_feature
93514..114081
/note="assembly_fragment:01026"
114182..126206
/note="assembly_fragment:01600
fragment_chain:2"

misc_feature
126307..144583
/note="assembly_fragment:01198
fragment_chain:2"

misc_feature
144684..147136
/note="assembly_fragment:01132
fragment_chain:2"

misc_feature
147237..164925
/note="assembly_fragment:01005
fragment_chain:2"

misc_feature
165026..169362
/note="assembly_fragment:01413
clone end:SP6
vector_side:right"

ORIGIN
Query Match 47.3% Score 512.2; DB 2; Length 169362;
Best Local Similarity 99.4%; Pred. No. 1.5e-106;
Matches 514; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 568 TATATCAGGAAACATGTTCAACCAATGCGCAAGTGGTGAACAGCAATGAACAGG 627
118303 TAAATCAGGAAACATGTTCAACCAATGCGCAAGTGGTGAACAGCAATGAACAGG 118362

QY 628 AATTATATGAGATGATGAATGTAAAGCCAGCCCAAGAAAGGGGGAGAGACATGTT 687
118363 AATTATATGAGATGATGAATGTAAAGCCAGCCCAAGAAAGGGGGAGAGACATGTT 118422

QY 688 TGATTCTCAGCCTGTTCCAAATTTGTGTAAAGACCTTTAACAAGTTGGCTGAATTTGG 747
118423 TGATTCTCAGCCTGTTCCAAATTTGTGTAAAGACCTTTAACAAGTTGGCTGAATTTGG 118482

Db 748 AGCAGAGTTCAAGAACATGAAGCAACATATPACAGAAATATTTCTTTACAGTGAAGACC 807
118483 AGCAGAGTTCAAGAACATGAAGCAACATATPACAGAAATATTTCTTTACAGTGAAGACC 118542

```
808 TACTTATCTGGGAATGAAACATCTGTTTTGGGCCAAGAAAGACAGCTTGCTT 867
118543 TACTTATCTGGGAATGAAACATCTGTTTTGGGCCAAGAAAGACAGCTTGCTT 118602
868 AGCCATTAATAAGATTTTATTAACCCCTTCAACACATTTGCCAATTAAGATTTCTGT 927
118603 AGCCATTAATAAGATTTTATTAACCCCTTCAACACATTTGCCAATTAAGATTTCTGT 118662
928 GAGTCTCTTGCAAAATTTTGATGAGTATGTCACAAACAGTCTTATTTGTTATAA 967
118663 GAGTCTCTTGCAAAATTTTGATGAGTATGTCACAAACAGTCTTATTTGTTATAA 118722
988 TTTTGAATATGTTGTTTATTAACCTATGAAATTCCTTTTATTAATAACATTAAGAAAT 1047
118723 TTTTGAATATGTTGTTTATTAACCTATGAAATTCCTTTTATTAATAACATTAAGAAAT 118782
1048 CCCTTACCTATCAGAAACAAACACTCTCTGTTTA 1084
118783 CCCTTACCTATCAGAAACAAACACTCTCTGTTTA 118819

SULT 14
359875/c 251187 bp DNA linear HTG 13-JUN-2001
CUS FINITION Homo sapiens chromosome 13 clone RP11-453N22, 53 unordered pieces.
CESSION AL359875
RSION AL359875.3 GI:9864226
WORDS HTG; HTGS PHASE1; HTGS _CANCELLED.
URCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Aug 22, 2000 this sequence version replaced gi:8694433.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
Project Information
Center project name: BA453N22
----- Summary Statistics
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 220597 bases at least Q40
Consensus quality: 234299 bases at least Q30
Consensus quality: 241103 bases at least Q20
Insert size: 245987; sum-of-contigs
Insert size: 190986; agarose-fp
Quality coverage: 2.52x in Q20 bases; sum-of-contigs Quality
Coverage: 3.92x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2130: contig of 2130 bp in length
* 2131 2230: gap of 100 bp
* 2231 5689: contig of 3459 bp in length
* 5690 5790: gap of 100 bp
* 5790 10046: contig of 4257 bp in length
* 10047 10146: gap of 100 bp
*
* 10147 12624: contig of 2478 bp in length
* 12625 12724: gap of 100 bp
* 12725 20993: contig of 8269 bp in length
* 20994 21093: gap of 100 bp
* 21094 28936: contig of 7843 bp in length
* 28937 29036: gap of 100 bp
* 29037 32230: contig of 3194 bp in length
* 32231 32330: gap of 100 bp
* 32331 35153: contig of 2823 bp in length
* 35154 35253: gap of 100 bp
* 35254 37649: contig of 2336 bp in length
* 37650 37750: gap of 100 bp
* 37750 40245: contig of 2496 bp in length
* 40246 40346: gap of 100 bp
* 40346 48420: contig of 8075 bp in length
* 48421 48520: gap of 100 bp
* 48521 52169: contig of 3649 bp in length
* 52170 52269: gap of 100 bp
* 52270 56443: contig of 4174 bp in length
* 56444 56543: gap of 100 bp
* 56544 63088: contig of 6545 bp in length
* 63089 63188: gap of 100 bp
* 63189 65281: contig of 2093 bp in length
* 65282 65381: gap of 100 bp
* 65382 67940: contig of 2559 bp in length
* 67941 68040: gap of 100 bp
* 68041 75029: contig of 6989 bp in length
* 75030 75129: gap of 100 bp
* 75130 85567: contig of 10438 bp in length
* 85568 85667: gap of 100 bp
* 85668 89150: contig of 3483 bp in length
* 89151 89250: gap of 100 bp
* 89251 93380: contig of 4120 bp in length
* 93381 93480: gap of 100 bp
* 93481 95619: contig of 2139 bp in length
* 95620 95720: gap of 100 bp
* 95720 100592: contig of 4872 bp in length
* 100593 100692: gap of 100 bp
* 100693 103959: contig of 3267 bp in length
* 103960 104059: gap of 100 bp
* 104060 110871: contig of 6812 bp in length
* 110872 110971: gap of 100 bp
* 110972 115555: contig of 4584 bp in length
* 115556 115655: gap of 100 bp
* 115656 118593: contig of 2938 bp in length
* 118594 118693: gap of 100 bp
* 118694 123530: contig of 4837 bp in length
* 123531 123630: gap of 100 bp
* 123631 127087: contig of 3457 bp in length
* 127088 127187: gap of 100 bp
* 127188 129684: contig of 2497 bp in length
* 129685 129784: gap of 100 bp
* 129785 132132: contig of 2348 bp in length
* 132133 132232: gap of 100 bp
* 132233 143792: contig of 11560 bp in length
* 143793 143892: gap of 100 bp
* 143893 148766: contig of 4874 bp in length
* 148767 148866: gap of 100 bp
* 148867 152258: contig of 3392 bp in length
* 152259 152358: gap of 100 bp
* 152359 156922: contig of 4564 bp in length
* 156923 157022: gap of 100 bp
* 157023 160153: contig of 3131 bp in length
* 160154 160253: gap of 100 bp
* 160254 164334: contig of 4081 bp in length
* 164335 164434: gap of 100 bp
* 164435 166000: contig of 2166 bp in length
* 166001 166700: gap of 100 bp
* 166701 171896: contig of 5196 bp in length
* 171897 171996: gap of 100 bp
* 171997 184915: contig of 12919 bp in length
* 184916 185015: gap of 100 bp
* 185016 187238: contig of 2223 bp in length
```

```

* 187239 187338: gap of 100 bp
* 187339 188602: config of 2266 bp in length
* 189603 189702: gap of 100 bp
* 189703 194463: config of 4761 bp in length
* 194464 194563: gap of 100 bp
* 194564 197682: config of 3119 bp in length
* 197683 197783: gap of 100 bp
* 197784 201788: config of 4006 bp in length
* 201789 204060: gap of 100 bp
* 204061 204160: config of 2172 bp in length
* 204161 206237: gap of 100 bp
* 206238 206337: config of 2077 bp in length
* 206338 210376: gap of 100 bp
* 210377 210477: config of 4039 bp in length
* 210478 225245: gap of 100 bp
* 225246 225345: config of 14769 bp in length
* 225346 230027: gap of 100 bp
* 230028 230127: config of 4682 bp in length
* 230128 23610: gap of 100 bp
* 23611 232710: config of 2483 bp in length
* 232711 234942: gap of 100 bp
* 234943 235042: config of 2232 bp in length
* 235043 240958: gap of 100 bp
* 240959 241058: config of 5916 bp in length
* 241059 251187: gap of 100 bp

```

```

FEATURES
Source
Location/Qualifiers
1..251187
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone_id="RP11-453N22"
1..251187
/location="RP11-453N22"

```

```

misc_feature /note="assembly fragment:00469"
fragment_chain:1
2231..5689
/note="assembly fragment:01219"
fragment_chain:1
5790..10046
/note="assembly fragment:00352"
fragment_chain:1
10147..13624
/note="assembly fragment:00547"
fragment_chain:2
12725..20993
/note="assembly fragment:00617"
fragment_chain:2
21094..28936
/note="assembly fragment:01248"
fragment_chain:2
29037..33230
/note="assembly fragment:00606"
fragment_chain:3
33231..35153
/note="assembly fragment:01016"
fragment_chain:3
35254..37649
/note="assembly fragment:01277"
fragment_chain:3
37750..40245
/note="assembly fragment:00920"
fragment_chain:4
40346..48420
/note="assembly fragment:01288"
fragment_chain:4
48521..52169
/note="assembly fragment:00615"
fragment_chain:4
52270..56443
/note="assembly fragment:00009"
fragment_chain:5

```

```

misc_feature 56544..63088
/note="assembly fragment:00448"
fragment_chain:5
63189..65281
/note="assembly fragment:00027"

Query Match
Best local similarity 99.4%, Pred. No. 1,7e-106;
Matches 514; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

misc_feature /note="assembly fragment:00027"

568 TATATCAGAAACATGTTCAACCAATGCGAAGTGGTGAACACAGACATGAACACAG 627
Db TAAATCAGAAACATGTTCAACCAATGCGAAGTGGTGAACACAGACATGAACACAG 194040

628 AATTTATATAGACATGATGATGAAGACACCCAGAAAGGGGCGAGACATGTT 687
Db TAAATCAGAAACATGTTCAACCAATGCGAAGTGGTGAACACAGACATGAACACAG 194040

194039 TAAATCAGAAACATGTTCAACCAATGCGAAGTGGTGAACACAGACATGTT 193980
Db TAAATCAGAAACATGTTCAACCAATGCGAAGTGGTGAACACAGACATGAACACAG 193980

688 TGATTCCTACACATGTTCCAAATTTGTTTAAAGACCTTTAAACAGTTGGCTGAATTTGG 747
Db TGATTCCTACACATGTTCCAAATTTGTTTAAAGACCTTTAAACAGTTGGCTGAATTTGG 193920

748 AGCAGAGTTCAAGAACATGAACCAACCTATACAGAAATATTTCTTACAGTGAAGAAC 807
Db AGCAGAGTTCAAGAACATGAACCAACCTATACAGAAATATTTCTTACAGTGAAGAAC 193860

808 TACTTATCTGGGAATGAACATCTGTTTGGGCGCAACAGAAACAAACATCTTGTT 867
Db TACTTATCTGGGAATGAACATCTGTTTGGGCGCAACAGAAACAAACATCTTGTT 193800

868 AGCCATTAAGATTTATTAACCCCTCAACACATTTGCCAATGAATTTCTGTT 927
Db AGCCATTAAGATTTATTAACCCCTCAACACATTTGCCAATGAATTTCTGTT 193740

928 GAGTCTCTGCAATTTTATGAGATGATGACCAACAGTCTATTTGTTATATA 987
Db GAGTCTCTGCAATTTTATGAGATGATGACCAACAGTCTATTTGTTATATA 193680

988 TTTTGAATATGTTTATTTTACCTATGAATTTCCCTTTATTTAAATATGATGAAGAAAT 1047
Db TTTTGAATATGTTTATTTTACCTATGAATTTCCCTTTATTTAAATATGATGAAGAAAT 193620

1048 CCCTTATCTATCAAGAAACAAACACCTCTGTTTA 1084
Db CCCTTATCTATCAAGAAACAAACACCTCTGTTTA 193563

193619 CCCTTATCTATCAAGAAACAAACACCTCTGTTTA 193563

RESULT 15
BD060465 506 bp DNA linear PAT 27-AUG-2002
LOCUS Secreted expressed sequence tags (ESTs).
DEFINITION BD060465.1 GI:22606071
ACCESSION BD060465.1
VERSION UP 2001518793-A/825.
KEYWORDS Zea mays
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 506)
REFERENCE
AUTHORS Jacobs, K., McCoy, J.M., Lavallie, E.R., Racine, L.A., Metberg, D.,
Treacy, M., Spaulding, V., and Agostino, M.J.
TITLE Secreted expressed sequence tags (ESTs)
JOURNAL Patent: JP 2001518793-A 825 16-OCT-2001;
GENETICS INSTITUTE INC
PN JP 2001518793-A/825
PD 16-OCT-2001
PF 10-APR-1998 JP 1998543070
PR 10-APR-1997 US 08/837312
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACINE, PI
DAVID METBERG
PT MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
CI N15/12, C12N5/10, C07K14/47, C12Q1/68, A61K38/17 CC Strandness:

```

Double;
CC Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
source 1..506
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"

ORIGIN

Query Match 37.0%; Score 401.4; DB 6; Length 506;
Best Local Similarity 99.8%; Pred. No. 1.1e-81;
Matches 402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	497	GCTGCTGCTTTTGGAGGAATTGATGATGTTCACTGGAAGAAAATGGACATTAGTT	556
DB	99	GCTGCTGCTTTTGGAGGAATTGATGATGTTCACTGGAAGAAAATGGACATTAGTT	158
QY	557	CAAGTAGCACTATATCAGAAACATGTCAACCAATGGCAAGTGGGTGAACAGGAC	616
DB	159	CAAGTAGCACTATATCAGAAACATGTCAACCAATGGCAAGTGGGTGAACAGGAC	218
QY	617	AATGAAACAGAAATTTATATGAGACATGGAATGTAAAAGCCAGCCAGAAAAGGGGCA	676
DB	219	AATGAAACAGAAATTTATATGAGACATGGAATGTAAAAGCCAGCCAGAAAAGGGGCA	278
QY	677	GAGACATGTTGATTCCTACAGACTGTTCCAAATTTGTGTTAAGACCTTTACAGATTG	736
DB	279	GAGACATGTTGATTCCTACAGACTGTTCCAAATTTGTGTTAAGACCTTTACAGATTG	338
QY	737	GCTGAATTTGAGCAGAGTTCAGAAACATAGAAACCACTATACAGAAATTTCTTTAC	796
DB	339	GCTGAATTTGAGCAGAGTTCAGAAACATAGAAACCACTATACAGAAATTTCTTTAC	398
QY	797	AGTGAAGAACTTACTATCTGGAATGAAACATCTGTTTTGGGCCAACAGAAAACAAG	856
DB	399	AGTGAAGAACTTACTATCTGGAATGAAACATCTGTTTTGGGCCAACAGAAAACAAG	458
QY	857	ACTCTGGTTTAGCCATAAAAGATTTTATTCACCCCTTCAAC	899
DB	459	ACTCTGGTTTAGCCATAAAAGATTTTATTCACCCCTTCAAC	501

Search completed: April 22, 2004, 02:57:20
Job time : 4425.69 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

3M nucleic - nucleic search, using sw model

Run on: April 22, 2004, 00:17:22 ; Search time 2854.26 Seconds

(without alignments)
9918.266 Million cell updates/sec

Title: US-10-010-050A-1_COPY_137_1084

Perfect score: 948

Sequence: 1 cgggctccgggcatccctc.....acaacactctgtttta 948

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 1491090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estov:*
5: em_estpl:*
6: em_estro:*
7: em_estro:*
8: em_estro:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vit:*
21: em_ges_fun:*
22: em_ges_man:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rtd:*
26: em_ges_phg:*
27: em_ges_vrl:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	864.8	91.2	1201 9	AL546472
2	691.8	72.0	13	BX31615
3	683.6	73.1	13	BX342662
4	674.2	71.1	9	AL571805

5	673	71.0	931	10	BE873363	BE873363	601450425
6	671.8	70.9	2356	11	AK043247	AK043247	Mus muscu
7	670.6	70.7	689	13	BU621797	BU621797	UI-H-FL1
8	670.2	70.7	2433	11	AK085741	AK085741	Mus muscu
9	668.6	70.5	2247	11	AK032293	AK032293	Mus muscu
10	645.4	68.1	908	13	BX370641	BX370641	BX370641
11	643.2	67.8	924	14	CA488543	CA488543	AGENCOURT
12	629.6	66.4	869	12	BI090566	BI090566	60285673
13	601	63.4	787	12	BI223533	BI223533	602941923
14	600	63.3	641	14	CB215456	CB215456	NI5C_HD06
15	562.4	59.3	1098	10	BF982158	BF982158	602308796
16	550.4	58.1	915	12	BI080303	BI080303	602876736
17	546.2	57.6	564	14	CB297292	CB297292	12B22045
18	542.6	57.2	755	13	BQ443453	BQ443453	UI-M-EV0-
19	522.8	55.1	974	13	BQ919385	BQ919385	AGENCOURT
20	507.8	53.6	676	12	BQ620113	BQ620113	602782325
21	506.2	53.4	938	13	BQ926170	BQ926170	AGENCOURT
22	501	52.8	607	28	AQ030925	AQ030925	CITBI-E1-
23	489.4	51.6	777	10	AM475730	AM475730	un71c09.Y
24	475	50.1	607	12	BM849262	BM849262	K-EST0129
25	474.6	50.1	828	12	BI156421	BI156421	602919530
26	468.8	49.5	872	12	BI734549	BI734549	60336375
27	468.2	49.4	833	14	CK128221	CK128221	AGENCOURT
28	458.4	48.4	796	12	BI182240	BI182240	un71c09.Y
29	455.6	48.1	853	12	BI456284	BI456284	603172673
30	447.2	47.2	599	13	BO417448	BO417448	IK38B05.Y
31	442.4	46.7	706	10	AM594284	AM594284	h579710.X
32	435.6	45.9	706	10	BF739857	BF739857	7039H03.X
33	435	45.9	516	14	CB455600	CB455600	712753.MA
34	430	45.4	633	12	BM740661	BM740661	K-EST0012
35	429	45.3	565	9	AU280382	AU280382	AV723592
36	423.8	44.7	690	14	CF724176	CF724176	un71c09.Y
37	421.2	44.4	854	12	BI082103	BI082103	602877265
38	399.6	42.2	854	12	BI082103	BI082103	602877265
39	395	41.7	791	12	BG619558	BG619558	602618951
40	385.8	40.7	661	13	BY734042	BY734042	BY734042
41	383.4	40.4	465	12	BI304716	BI304716	AR064C061
42	382	40.3	825	13	BU633296	BU633296	un71c09.Y
43	380.6	40.1	767	9	AJ451053	AJ451053	AL451053
44	379.2	40.0	915	13	BX757396	BX757396	BX757396
45	377.6	39.8	3269	11	AK030852	AK030852	Mus muscu

ALIGNMENTS

RESULT 1
AL546472
LOCUS
DEFINITION
AL546472 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
clone CS0D1030Y01 5-PRIME, mRNA sequence.
ACCESSION
AL546472 GI:31268306
VERSION
AL546472.2
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12679620.
COMMENT
Contact: Genoscope - Centre National de Sequencage
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1030C0E01Q1&cluster=7238.f. Contact :
Feng Liang Email : liang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0D1030CE01Qp1.
 Location/Qualifiers
 1. 1201

FEATURES
 source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1030Y01"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 91.2%; Score 864.8; DB 9; Length 1201;
 Best Local Similarity 95.4%; Pred. No. 6.5e-186;
 Matches 904; Conservative 17; Mismatches 19; Indels 8; Gaps 3;

```

Qy 1 CGGGTCTCGGGGATCCCTCCCGGGCCCACTGGCCGGCCCTTACAGAGGCTTGACTTC 60
Db 153 CGGGTCTCGGGGATCCCTCCCGGGCCCACTGGCCGGCCCTTACAGAGGCTTGACTTC 212
Qy 61 CGTCCAAAACCTGATCCTTATTTGTCAAGCTAAGTACTTTCTGTCCAACCTGGCTCACCT 120
Db 213 CGTCCAAAACCTGATCCTTATTTGTCAAGCTAAGTACTTTCTGTCCAACCTGGCTCACCT 272
Qy 121 ATCCCAAGTTATGAGAGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
Db 273 ATCCCAAGTTATGAGAGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 332
Qy 181 GAATTTAAATATGAGAGCTCTCGGGAACCTGGAATAATATGATGATGATGATGATGATG 240
Db 333 GAATTTAAATATGAGAGCTCTCGGGAACCTGGAATAATATGATGATGATGATGATGATG 392
Qy 241 AGAAGTACATTAAGTACGAGAGAGTACACATGAGAGAGTATGAGAGAGTATGAGAGAGT 300
Db 393 AGAAGTACATTAAGTACGAGAGAGTACACATGAGAGAGTATGAGAGAGTATGAGAGAGT 452
Qy 301 AACTGTACATTTCCCACTCTCGAAGTGAATGATGATGATGATGATGATGATGATGATG 350
Db 453 AACTGTACATTTCCCACTCTCGAAGTGAATGATGATGATGATGATGATGATGATGATG 512
Qy 361 GGTGCTGCTTTTGGAGGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
Db 513 GGTGCTGCTTTTGGAGGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 572
Qy 421 CAAGTAGCACTATATCAAGAAACATGTTCAACCAATGGCAAGTGGGTGAAACAGAGAC 480
Db 573 CAAGTAGCACTATATCAAGAAACATGTTCAACCAATGGCAAGTGGGTGAAACAGAGAC 632
Qy 481 AATGAAACAGAAATTTATATGAGCATGAGATGTAAGAGCCAGCCAGAAAGGGGGGCA 540
Db 633 AATGAAACAGAAATTTATATGAGCATGAGATGTAAGAGCCAGCCAGAAAGGGGGGCA 692
Qy 541 GAGACATGGTGGATTCCTACAGACTGTTCCAAATTTGTTTGAAGACCTTTAACAAGTTG 600
Db 693 GAGACATGGTGGATTCCTACAGACTGTTCCAAATTTGTTTGAAGACCTTTAACAAGTTG 752
Qy 601 GCTGAATTTGAGAGAGTTCAGAAACATGAAACCAATATCAAGATATTTCTTTAC 650
Db 753 GCTGAATTTGAGAGAGTTCAGAAACATGAAACCAATATCAAGATATTTCTTTAC 812
Qy 661 AATGAGAACTTACTATCTGGGAAATGAAACATGTTTGGGGCCCAAGAGAAACAG 720
Db 813 AATGAGAACTTACTATCTGGGAAATGAAACATGTTTGGGGCCCAAGAGAAACAG 872
Qy 721 ACTCTGGTTTGAACATAAAGATTTTATTTACCCCTCAACACCATTTGGCACTATAA 780
Db 873 ACTCTGGTTTGAACATAAAGATTTTATTTACCCCTCAACACCATTTGGCACTATAA 932
Qy 781 GAATTTCTGTTGAGTCTCTTGAATTTTGTGATGAGATGATGATGATGATGATGATG 840
  
```

```

Db 993 GAATTTCTGTTGAGTCTCTTGAATTTTGTGATGAGATGATGATGATGATGATGATGATG 992
Qy 841 TTGTTTATATTTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Db 993 TTGTTTATATTTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1047
Qy 901 TATGAGAAATCCCTTATCTATGAGAAACCAACCTCTGGTTTGA 948
Db 1048 TATGAGAAATCCCTTATCTATGAGAAACCAACCTCTGGTTTGA 1092
  
```

RESULT 2

EX331615 923 bp mRNA linear EST 01-MAY-2003
 DEFINITION BX331615 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
 cDNA clone CS0D1009Y001 5-PRIME, mRNA sequence.
 ACCESSION BX331615
 VERSION BX331615.1 GI:30310073
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLES Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 7238.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0D1009A01Qp1&cluster=7238.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0D1009A01Qp1.

FEATURES

source

Location/Qualifiers
 1. 923
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1009Y001"
 /tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
 /clone_id="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 73.0%; Score 691.8; DB 13; Length 923;
 Best Local Similarity 96.5%; Pred. No. 1.2e-146;
 Matches 690; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

```

Qy 1 CGGGTCTCGGGGATCCCTCCCGGGCCCACTGGCCGGCCCTTACAGAGGCTTGACTTC 60
Db 208 CGGGTCTCGGGGATCCCTCCCGGGCCCACTGGCCGGCCCTTACAGAGGCTTGACTTC 267
Qy 61 CGTCCAAAACCTGATCCTTATTTGTCAAGCTAAGTACTTTCTGTCCAACCTGGCTCACCT 120
Db 268 CGTCCAAAACCTGATCCTTATTTGTCAAGCTAAGTACTTTCTGTCCAACCTGGCTCACCT 327
Qy 121 ATCCCAAGTTATGAGAGGGTATGATGATGATGATGATGATGATGATGATGATGATGATG 180
Db 328 ATCCCAAGTTATGAGAGGGTATGATGATGATGATGATGATGATGATGATGATGATGATG 387
Qy 181 GAATTTAAATATGAGAGCTCTCGGGAACCTGGAATAATATGATGATGATGATGATGATG 240
Db 388 GAATTTAAATATGAGAGCTCTCGGGAACCTGGAATAATATGATGATGATGATGATGATG 447
  
```

	ly	241	AACAATCATATTAACTGGCAAGAACATAACAAATGGAATGTAATCAACTTTTCCAACTTGCC	300
	lb	448	ACAGATCATTTAATCTGGAGAAACATTAAACATGGATATGTATBAATCTTTTCACACTTGCG	507
	ly	301	AACTGATCAATTTCCCACATCTCCGACTGAAATGGATGCCCTTTCTGTGTATCAAGGC	360
	lb	508	AANTGTANATTTCCCGCCTCTCCGNNCTGNATGGRTSSCCTTTCTGTGATATCAAGG	567
	ly	361	GCTGCTCTGTTTTTTGAGGGAAATTGATGATGTTCACTGGAAGGAAAATGGACATTAGTT	420
	ly	568	GSATGCTCTTTTTTGRGGGAATGTGTGATGTTTCACTGGAAGGAAAATGGACATTAGTT	627
	ly	421	CAGATGAGAATAATATATGAGAAACATGTCACCAAAATGGCAAAATGGGTGAACAGAC	480
	lb	628	CAGATGAGACATATATCACGAGAAACATGTTCAACCMAATGGCAAAATGGGTGAACAGAC	687
	ly	481	AATGAACAGAAATTTATATGAGACATGGAATGTAAATGAAGCCAGCCAGAAAAAGGGGCA	540
	ly	688	AATGAACAGAAATTTATATGAGACATGGAATGTAAATGAAGCCAGCCAGAAAAAGGGGCA	747
	ly	541	GAGACATGTTGATTTCCATAGACTGTCCAAATTTGTGTAGAGACTTTAACAAAGTG	600
	ly	748	GAGACATGTTGATTTCCATAGACTGTCCAAATTTGTGTAGAGACTTTAACAAAGTG	807
	ly	601	GCTGAATTTGGAGCAGAGTTCAAGAAATAGAAACCAATACAGAAATATTTCTTAG	660
	lb	808	GCTGAATTTGGAGCAGAGTTCAAGAAATAGAAACCAATACAGAAATATTTCTTAG	867
	ly	661	AGTGAAGAACCTTATCTTGCGGAATGAACATCTGTTTTTGGGCCAACAGAA	715
	lb	868	AGTGAAGAACCTTATCTTGCGGAATGAACATCTGTTTTTGGGCCAACAGRAA	922
RESULT 3				
LOCUS	BX342662	1077 bp	mRNA	linear EST 02-MAY-2003
DEFINITION	BX342662 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED			
ACCESSION	BX342662			Homo sapiens cDNA clone CSODL004YP03 5-PRIME, mRNA sequence.
VERSION	BX342662.1	GI:30338103		
KEYWORDS	EST.			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens (human)			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 1077)			
AUTHORS	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Genoscope			
	Genoscope - Centre National de Sequencage			
	BP 191 91006 Evry cedex - France			
	Email: seq@genoscope.cns.fr, Web : www.genoscope.cns.fr			
	Library was constructed by Life Technologies, a division of			
	Invitrogen. This sequence belongs to sequence cluster 7238.f For			
	more information about this cluster, see			
	http://www.genoscope.cns.fr/			
	cgi-bin/cluster.cgi?seq=CSODL004CH02QPl&cluster=7238.f. Contact :			
	Feng Liang Email : fliang@lifetech.com URL :			
	http://fulllength.invitrogen.com/ Invitrogen Corporation 1600			
	Paraday Avenue Genoscope sequence ID : CSODL004CH02QPl.			
FEATURES	Location/Qualifiers			
source	1..1077			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="CSODL004YP03"			
	/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"			
	/cote_line="RAMOS CELL LINE"			
	/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT			
	25-NORMALIZED"			
	/note="1st strand cDNA was primed with a NotI-oligo(dT)"			

ORIGIN

primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match	72.1%;	Score 683.6;	DB 13;	Length 1077;
Best Local Similarity	88.8%;	Pred. No. 8.8e-145;		
Matches 700;	Conservative 49;	Mismatches 34;	Indels 5;	Gaps 2;

[illegible]

KEYWORDS	EST.
SOURCE	
ORGANISM	
	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 1201)
AUTHORS	Li, M. B., Gruber, C., Jesse, J. and Polyes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	unpublished (2001)
COMMENT	On Feb 16, 2001 this sequence version replaced gi:12929467.

Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f. For
more information about this cluster, see
[http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1030CE01NP1&cluster=7238.f](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1030CE01NP1&cluster=7238.f). Contact :
Feng Liang Email: fliang@lifetech.com URL: <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0D1030CE01NP1.
Location/Qualifiers

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOD1030Y701"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_1b="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime and enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

Query Match	71.1%	Score 674.2	DB 9	Length 1201
Best Local Similarity	97.8%	Pred. No. 1.2e-142		
Matches 704	Conservative 9	Mismatches 4	Indels 3	Gaps 3
QY	230	CCATTGATTCAGAGTACATTAACTGCGAGAACTACACATGGAATGATGATGACTTT	289	
Db	1075	SCCATKGATTCAGAGTACATTAACTGCGAA-RAMTCAMAAATGAATGGTATG-AMTTT	1011	
QY	290	TCCACTTGGCACTGTACATTTCCCATCTCCGACTGAAATGATGCCCTTTCTGCT	349	
Db	1017	TCCAACTTGGCAACTGTACATTTCCCATCTCCGACTGAAATGATGCCCTTTCTGCT	958	
QY	350	GTAAACAAGGGCGCTGCTTTTGTAGGG-ATTGATGATGTACATGGAAGGAAT	408	
Db	957	GTAAACAAGGGCGCTGCTTTTGTAGGAAATGATGATGTACATGGAAGGAAT	898	
QY	409	GGGACATTAGTTCAAGTAGCAACTATATCAGAGAAACATGTTCAACCAATGGCAAGTGG	468	
Db	897	GGGACATTAGTTCAAGTAGCAACTATATCAGAGAAACATGTTCAACCAATGGCAAGTGG	838	
QY	469	GTGAACAAGGACATGAAACAGAAATTATATGACATGGAATGAAAGGCCAGCCCA	528	
Db	837	GTGAACAAGGACCAATGAAACAGAAATTATATGACATGGAATGAAAGGCCAGCCCA	778	
QY	529	GAAAGGGGGGCGAGAGACATGTTGATTTCTTACACATGTTCCAAATTTGTGTTAAGACC	568	
Db	777	GAAAGGGGGGCGAGACATGTTGATTTCTTACACATGTTCCAAATTTGTGTTAAGACC	718	
QY	589	TTTAACAAGTTGGCTGAATTTGGAGCGAGTTTCAAGAACATGAAACCAACTATCAAGA	648	
Db	717	TTTAACAAGTTGGCTGAATTTGGAGCGAGTTTCAAGAACATGAAACCAACTATCAAGA	658	
QY	649	ATAATTTCTTTACAGTGAAGAACCTACTATCTGGAATGAAACATCTGTTTGGGGCA	708	
Db	657	ATAATTTCTTTACAGTGAAGAACCTACTATCTGGAATGAAACATCTGTTTGGGGCA	558	
QY	709	ACAGGAAACAAGACTCTTGTTAGCCATATAAAGATTTTATTAACCTTCAACACCAT	768	

Db 597 ACAGAAACGAAGACTCTTGCTTTAGCCATAAAAAGATTTTATTACCCTCTCAAAACCACT 538

QY 769 TTGCCAACTAAGAATTTCTGTGAGTCTCTGCAAATTTTGTAGTCAGTGATTTGTGCAC 828

Db 537 TTGCCAACTAAGAATTTCTGTGAGTCTCTGCAAATTTTGTAGTCAGTGATTTGTGCAC 478

QY 829 AAACAGTTCTATTTTGTGTTTATTAATTTTGAATTTGGTITTTACCTATGAAATTCCTTT 888

Db 477 AAACAGTTCTATTTGTTTATTAATTTTGAATTTGGTITTTACCTATGAAATTCCTTT 418

QY 889 ATTAAAAATACATATGAAGAATTCCTTTACCTATCAGAAAACAAACACTCTCTGGTTTA 948

Db 417 ATTAAAAATACATATGAAGAATTCCTTTACCTATCAGAAAACAAACACTCTCTGGTTTA 358

RESULT 5	BE873363	931 bp	mRNA	linear	EST 20-OCT-2000
LOCUS	BE873363				
DEFINITION	BE873363	601450425F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3854119 5',			
LOCUS	BE873363	mRNA sequence.			
ACCESSION	BE873363				
VERSION	BE873363.1	GI:10322139			
KEYWORDS	EST.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				
	Homo sapiens				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 931)	NIH-MGC	http://mgc.nci.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)		Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINT at:
<http://image.llnl.gov>
plate: L1AM9579 row: b column: 08
High quality sequence stop: 662.

FEATURES
SOURCE

```

/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3854119"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH_MGC_65"
/note="organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 Kb. Library constructed by Life
Technologies."

```

ORIGIN

	Query Match	71.0%	Score 673	DB 10	Length 931
	Best Local Similarity	93.0%	Pred. No. 2.3e-142		
	Matches 727	Conservative 0	Mismatches 51	Indels 4	Gaps 2
QY	138	TGATGATGACATTGAACTTTTCGATATACAAAGCCCGAGATGGGAATTTAATATGAGAGA	197		
Db	1	TGATGATGACATTGAACTTTTCGATATACAAAGCCCGAGATGGGAATTTAATATGAGAGA	60		
QY	198	CTCTCTGGGACACTTGAAAATTATGATGATGCCATTGGATTCAGAACTACATTACTGG	257		
Db	61	CTCTCTGGGACACTTGAAAATTATGATGATGCCATTGGATTCAGAACTACATTACTGG	120		
QY	258	CAGAACTACAAATGGAATGGAATGGAATCTTTCCAACTTGGCAACGTACATTCCCA	317		
Db	121	CAGAACTACAAATGGAATGGAATGGAATCTTTCCAACTTGGCAACGTACATTCCCA	180		

```

2y 318 TCTCCGACCTGAATGATGCGCCCTTTCTGTGTTAATCAAGGCGCTGCTTTTGA 377
2b 181 TCTCCGACCTGAATGATGCGCCCTTTCTGTGTTAATCAAGGCGCTGCTTTTGA 240
2y 378 GGGAAATGATGATGTTCACTGGAGAAATGGGACATTAGTTCAAGTAGCACTATATC 437
2b 241 GGGAAATGATGATGTTCACTGGAGAAATGGGACATTAGTTCAAGTAGCACTATATC 300
2y 438 AGGAAACATGTTCAACCAATGGCAAAAGTGGGTAAACAGGACATGAAACAGAAATTTA 457
2b 301 AGGAAACATGTTCAACCAATGGCAAAAGTGGGTAAACAGGACATGAAACAGAAATTTA 360
2y 498 TTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 557
2b 361 TTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
2y 558 CTACGACTGTTCCAAATTTGTTTAAAGACCTTTTAAAGATTTGCTGATTTTGAAGACA 617
2b 421 CTACGACTGTTCCAAATTTGTTTAAAGACCTTTTAAAGATTTGCTGATTTTGAAGACA 480
2y 618 GTTCAAGACATGAGAAACCAATCTTACAGAAATTTCTTTACGTGAGACCTACTTA 677
2b 481 GTTCAAGACATGAGAAACCAATCTTACAGAAATTTCTTTACGTGAGACCTACTTA 540
2y 678 TCTGGGAAATGAACATCTGTTTGGGCGCAAGAGAAACAGACTGTTGGTTAGGCAT 737
2b 541 TCTGGGAAATGAACATCTGTTTGGGCGCAAGAGAAACAGACTGTTGGTTAGGCAT 600
2y 738 AAAAAGATTTTATTTACCCCTTCAACACACATTTGCCAATTAAGAAATTTCTGTTAGCT 797
2b 601 AAAAAGATTTTATTTACCCCTTCAACACACATTTGCCAATTAAGAAATTTCTGTTAGCT 660
2y 798 CTTCGAATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 857
2b 661 CTTCGAATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 716
2y 858 ATATGATTTTATTTACCTTGAATTTCCCTTTTATTAATAATGATGATGATGATGATG 917
2b 717 ATATGATTTTATTTACCTTGAATTTCCCTTTTATTAATAATGATGATGATGATGATG 776
2y 918 AC 919
2b 777 TC 778

```

RESULT 6
AK043247
LOCUS
DEFINITION Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone: A730075N08 product: similar to CEROID-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CNS PROTEIN) [Homo sapiens], full insert sequence.
ACCESSION AK043247
VERSION AK043247.1 GI:26089577
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, N., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Mech. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, N., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374

PUBMED 11042159
REFERENCE
AUTHORS 3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kikunaka, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE
AUTHORS 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE
AUTHORS 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE
AUTHORS 6 (bases 1 to 2356)
TITLE Submitted (16-JUL-2001) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
FEATURES
source 1. .2356
organism "Mus musculus"
mol_type "mRNA"
strain "C57BL/6J"
db_xref "FANTOM:DB:A730075N08"
db_xref "MGI:2408405"
db_xref "taxon:10090"
clone "A730075N08"
tissue type="cerebellum"
clone lib="RIKEN full-length enriched mouse cDNA library"
dev stage="7 days neonate"
misc_feature 1. .2356
note "similar to CEROID-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CNS PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence: FASTA, 73.9%ID, 86.7%length, match=1041)"
ORIGIN
Query Match 70.9%; Score 671.8; DB 11; Length 2356;

Best Local Similarity 81.8%; Pred. No. 3.9e-142;
Matches 775; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

```

QY 1 CGGGTCTCGAGCATCCCTCCCGGCGCACTGCGCGGCTGCTTCAAGCGCTTACTTC 63
DB 130 CTGGGCGCGTCCCGACGCTCTGGGCAACGCTGGCCGGTCCCTTAAGCGCTTCTTTC 139
QY 61 CGTCCAAAACCTGATCTTATTTGCAAGTAAGTACTTTCTGTCACATGGCTCACCT 120
DB 130 CCGCCGAAGACAGATCCCTACTGCTCAAGCTAAGTACTTTCTGCTCAACGGCTCGCC 249
QY 121 ATCCCAAGTTATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 130
DB 250 ATCCCAAGTTATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 309
QY 181 GAATTTAATATGAGACCTCTGGGACATGAAATATGATGATGATGATGATGATGATG 240
DB 310 GAATTTAATATGAGACCTCTGGGACATGAAATATGATGATGATGATGATGATGATG 369
QY 241 AGAAGTACATTAAGTGGAGAGAACTACAAGTGAATGATGATGATGATGATGATGATG 300
DB 370 AGAAGTACATTAAGTGGAGAGAACTACAAGTGAATGATGATGATGATGATGATGATG 429
QY 301 AACTGTATATTTCCCGCATCTCGACACTGAATGATGATGATGATGATGATGATGATG 360
DB 430 AACTGTATATTTCCCGCATCTCGACACTGAATGATGATGATGATGATGATGATGATG 489
QY 361 GCTGCTCTCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 480 GCAAGCTCTCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 549
QY 421 CAAGTACATTAATGAGAAACATGTTCAACCAATGAGAAATGGAGTGAACAGAGAC 480
DB 550 GTCGTGCAACATATCCGAAACATTTAACAAGAGGCGGATGATGATGATGATGATGATG 609
QY 481 AATGAAACAGAAATTTATGAGACATGATGATGATGATGATGATGATGATGATGATG 540
DB 610 AATGAAACAGAAATTTATGAGACATGATGATGATGATGATGATGATGATGATGATG 669
QY 541 GAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
DB 670 CAGAGTGTGCTGAGTCTTACGACCTGTTGCAATTTTGTCTTAAGACATTAAGAAATG 729
QY 601 GCTGAATTTGAGACAGAGTCAAGAACTATGAAACCACTATGACAAATTTCTTTAC 660
DB 730 GCTGAATTTGAGACAGAGTCAAGAACTATGAAACCACTATGACAAATTTCTTTAC 789
QY 661 AGTGAAGAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
DB 790 AGTGAAGAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 849
QY 721 ACTCTTGTGTTAGCCATTAAGAAATTTATTTACCCCTCAACCAATTTGCCAATGAA 780
DB 850 ACTCTTGTGTTAGCCATTAAGAAATTTATTTAGCCCTTCAACCAATTTGCCAATGAA 909
QY 781 GAATTTCTGTTGAGTCTCTTCAATTTTGAATGATGATGATGATGATGATGATGATGAT 840
DB 910 GAATTTCTGTTGAGTCTCTTCAATTTTGAATGATGATGATGATGATGATGATGATGAT 969
QY 841 TTGTTTATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
DB 970 TTGTTTATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1029
QY 901 TATGAAGAAATCCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 947
DB 1030 TATGAAGAAATCCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1076

```

RESULT 7

BU621797/c

LOCUS BU621797 689 bp mRNA linear EST 23-SEP-2002
DEFINITION UI-H-FL1-bgd-i-07-0-UI-s1 NCI CGAP_Fl1 Homo sapiens cDNA clone
UI-H-FL1-bgd-i-07-0-UI-3', mRNA sequence.

ACCESSION BU621797
VERSION BU621797.1 GI:23288012
KEYWORDS EST
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 689)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Email: cgapdb-remail.nih.gov

Tissue Procurement: James Martin
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained

from Dr. M. Bento Soares, bento-soares@uiowa.edu

Seq primer: M13 FORWARD

POLYA=yes.

FEATURES

Source

Location/Qualifiers

1..689

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-H-FL1-bgd-i-07-0-UI"

/tissue_type="Cell lines"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies)"

/note="Organ: Chondrosarcoma; Vector: pRT3-Pac

(pharmacia) with a modified polylinker; Site 1: EcoR I;

Site 2: Not I; NCI CGAP_Fl1 is a normalized cDNA library

derived from a pool of mRNA obtained from 4 cell lines

from grade III chondrosarcoma tissues. The library was

constructed according to Bonaldi, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA

synthesis was primed with an oligo-dT primer containing a

Not I site. Double stranded cDNA was ligated to an EcoR I

adaptor, digested with Not I, and cloned directionally

into pRT3-Pac vector. The oligonucleotide used to prime

the synthesis of first-strand cDNA contains a library tag

sequence that is located between the Not I site and the

(dT)18 tail. The sequence tag for this library is

GAGTGGGTC. The cell lines were provided by Dr. James

Martin from the University of Iowa.

TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix

TAG LIB=UI-H-FL1

TAG_SEQ=GAGTGGGTCG

ORIGIN

Query Match

Best Local Similarity 70.7%; Score 670.6; DB 13; Length 689;

Matches 673; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 234 TGGATTGAGAAAGTACATTAAGTGGCAAGAACTACACAAATGGAATGATGAACTTTTCCA 293
DB 689 TGGATTGAGAAAGTACATTAAGTGGCAAGAACTACACAAATGGAATGATGAACTTTTCCA 630
QY 294 ACTTGGCACTGATACATTTCCCATCTCCGACCTGAATGATGATGATGATGATGATGATG 353
DB 629 ACTTGGCACTGATACATTTCCCATCTCCGACCTGAATGATGATGATGATGATGATGATG 570
QY 354 TCAAGGCGCTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATG 413
DB 569 TCAAGGCGCTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATG 510
QY 414 ATTAGTTCAGTAGCACTATATCAGGAACATGTTCAACCAATGCAAAAGTGGGTGAA 473
DB 509 ATTAGTTCAGTAGCACTATATCAGGAACATGTTCAACCAATGCAAAAGTGGGTGAA 450

```

474 ACAGACATGAAACAGAAATTTATATGACATGGAATGTAAAGCCAGCCAGAAAA 533
 b 449 ACAGACATGAAACAGAAATTTATATGACATGGAATGTAAAGCCAGCCAGAAAA 393
 Y 534 GGGGGCAGACATGGTTGATTCCTACAGACTGTTCCAAATTTGTTAGACCTTTAA 593
 b 389 GGGGGCAGACATGGTTGATTCCTACAGACTGTTCCAAATTTGTTAGACCTTTAA 333
 Y 594 CAAAGTGGCTGAATTTGGAGCAGAGTTCAAGAACATGAAACCACTATACAGAAATTT 653
 b 329 CAAAGTGGCTGAATTTGGAGCAGAGTTCAAGAACATGAAACCACTATACAGAAATTT 273
 Y 654 TCTTTACAGAGGAGAACCTTATCTGGGAATGAAACATCTGTTTGGGCCAACAG 713
 b 269 TCTTTACAGAGGAGAACCTTATCTGGGAATGAAACATCTGTTTGGGCCAACAG 213
 Y 714 AAACAGACTCTGTTGATTCCTACAGACTGTTCCAAATTTGTTAGACCTTTAA 773
 b 209 AAACAGACTCTGTTGATTCCTACAGACTGTTCCAAATTTGTTAGACCTTTAA 153
 Y 774 AACTAAGAAATTTCTGTTAGTCTCTGCAATTTTGAATGCAAGTATGTCACAAACA 833
 b 149 AACTAAGAAATTTCTGTTAGTCTCTGCAATTTTGAATGCAAGTATGTCACAAACA 90
 Y 834 GTTCTATTTGTTTAAATTTGAATATGTTTGAATGTTTGAATGTTTGAATGTTTGA 893
 b 89 GTTCTATTTGTTTAAATTTGAATATGTTTGAATGTTTGAATGTTTGAATGTTTGA 30
 Y 894 AATTAACATATGAAAGAA 910
 b 29 AATTAACATATGAAAGAA 13

RESULT 8
 AK085741 2433 bp mRNA linear HTC 20-SBP-2033
 ACUS AK085741
 DEFINITION Mus musculus 10 days lactation, adult female mammary gland cDNA,
 RIKEN full-length enriched library, clone:DJ30033P03
 Product: similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5
 PROTEIN) [Homo sapiens], full insert sequence.
 ACCESSION AK085741
 VERSION AK085741.1 GI:26102938
 EMBL AK085741.1
 SOURCE HTC; CAP trapper.
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carinci, P. and Hayashizaki, Y.
 TITLE High efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 AUTHORS Carinci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Komoto, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
 Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN Integrated Sequence Analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913

PUBMED 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 2433)
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirozane, T.,
 Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Konda, M.,
 Koyama, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Morita, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ono, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-Apr-2002) Yoshinori Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan [E-mail: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216]
 COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 FEATURES
 source location/Qualifiers
 1..2433
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:DJ30033P03"
 /db_xref="MGI:2422937"
 /db_xref="taxon:10090"
 /clone="DJ30033P03"
 /sex="female"
 /tissue_type="mammary gland"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days lactation, adult"
 1..2433
 /note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5
 (CLN5 PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence:
 FASTA, 73.9%ID, 86.7%length, match=1041)"

ORIGIN
 misc_feature
 1..2433
 Query Match 70.7%; Score 670.2; DB 11; Length 2433;
 Best Local Similarity 81.7%; Pred. No. 8.9e-142;
 Matches 774; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

Y 1 CGGGTCTGGGGGATCCCTCCGGGCGCACTGGCGGCTCCCTACAGGCTTACTGCTC 60
 Db 207 CTGGGCGGCTCCCGACGCTGCGGCAAGCGTGGCGGCTCCCTACAGGCTTCTTTT 266
 Y 61 CGTCAAAACCTGATCTTATTTGCAAGTAAAGTACTTTCTGTCGAACCTGGCTACCT 120
 Db 267 CGTCCGAAGACAGATCCCTACGTCAAGTAAAGTACTTTCTGTCCTACCGGCTGGCC 326
 Y 121 ATCCAGTTATGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 180

CDS

```

/sex="male"
/tissue_type="olfactory brain"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
<1..971
/notes="unnamed protein product; putative
similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CNS
PROTEIN) [Homo sapiens] (SWISSPROT|075503, evidence:
FASTA, 73.9%ID, 86.7%length, match=1041)"
FASTA_start=3
/codon_start=3
/protein_id="BAC27797.1"
/db_xref="GI:26328113"
/translation="LGLATILGASPTSGORPVYKHSFRPKTDYCAKATYECPLG
SPRPMDNDVIEVLRIQAPLWERYGCDLGEPLMDADYGRSTLGRKATYEMKEL
FOLANCTPPHRRPKSAPFMNGSAACTFEGSIDDKMKENGTSVATISGNTKXA
EWMKQDNBTGIIYETWTVRAGPQQAQWESIDCSNPLRTYKLAEPTEERKJL
NYTKIFLYSGEPIYLGNETSI FGPKNKTLALAKKFGPFRPLSTDKDLMNFKLF
DVIHHRQFLYLFNFYFWFLPMKPPFKITVEETPLTRHTFTDL"

```

ORIGIN

Query Match 70.5%; Score 668.6; DB 11; Length 2247;

Best Local Similarity 81.6%; Pred. No. 2.1e-141;

Matches 773; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

```

1 CGGGCTCGGGGCAATCCCTCCCGGCGCACTGCGCGGCTTCAAGCGCTTGACTTC 60
21 CTGGGCGCGTCCCGGCACTGCGGCAAGCGTGGCGGCTTCAAGCACTTCTTTTC 80
61 CGTCCAAACCTGATCTTATTTGCTAAGTATATCTTCTGCTCAACTGGCTCACT 120
81 CGTCCGAAGACAGATCCCTACTGTCAAGCTAATATCTTCTGCTCAAGCGCTTCC 140
121 ATCCCACTTATGAGGGTATGATGATGATGATGATGATGATGATGATGATGATG 160
141 ATCCCACTTATGAGGACATGATGATGATGATGATGATGATGATGATGATGATG 200
181 GAATTTAATATGAGGACCTCTCGGACACTGCGGAAATTTGATGATGATGATGATG 240
201 GAATTTAATATGAGGACCTCTCGGACACTTAACTTAACTTAACTTAACTTAACT 260
241 AGAATGATCATTTAATGAGGACATGATGATGATGATGATGATGATGATGATGATG 300
261 AGAGGACACATGACAGGACAAAGATGATGATGATGATGATGATGATGATGATGATG 320
301 AACTGTACATTTTCCCACTTCCGACCTGGAATGATGATGATGATGATGATGATGATG 360
321 AACTGTACATTTTCCCACTTCCGACCTGGAATGATGATGATGATGATGATGATGATG 380
361 GCTGCGCTGCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATG 420
381 GCACTGCTGCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATG 440
421 CAATGATCAATATATGAGGACATGATGATGATGATGATGATGATGATGATGATGATG 460
441 GTGCTGCAACCAATATCGGAAACATTTAAAGAAATGAGGATGATGATGATGATGATG 500
481 AATGAAACAGGAATTTTATGAGACATGATGATGATGATGATGATGATGATGATGATG 540
501 AATGAAACAGGAATTTTATGAGACATGATGATGATGATGATGATGATGATGATGATG 560
541 GAGACATGCTTTGATGCTGACGCTGTTCAAAATTTGATGATGATGATGATGATGATG 600
561 CAGACGCTGCTTTGATGCTGACGCTGTTCAAAATTTGATGATGATGATGATGATGATG 620
601 GCTGAATTTGAGACAGATTCAGAAACATGAAACCAACTATATCAAGATATTTCTTAC 660
621 GCTGAATTTGAGACAGATTCAGAAACATGAAACCAACTATATCAAGATATTTCTTAC 680
661 AGTGAAGAACTTACTTATCTGGGAAATGAAACATCTTTTGGGCGCAACAGAAACAG 720
681 AGTGAAGAACTTACTTATCTGGGAAATGAAACATCTTTTGGGCGCAACAGAAACAG 740
721 ACTCTGTTAGCATATAAAAGATTTTATATCCCTTCAACCAACTTTGCCAACTAA 780

```

```

Db 741 ACTCTGCTTTGAGCATATAAAAGATTTTATATCCCTTCAACCAACTTTGCCAACTAA 800
Qy 781 GAAATTTCTGTTGAGCTCTTTCGAAATTTTATGATGATGATGATGATGATGATGATGATGATG 840
Db 801 GATTTCTGATGATGATTTCTTGAATAATTTTATGATGATGATGATGATGATGATGATGATGATG 860
Qy 841 TTGTTTATATATTTTGAATTTGTTTATGTTTATGATGATGATGATGATGATGATGATGATGATG 900
Db 861 TTGTTTATATATTTTGAATTTGTTTATGTTTATGATGATGATGATGATGATGATGATGATGATG 920
Qy 901 TATGAAGAAATCCCTTACTATCAGAAACAACTCTCTGATTT 947
Db 921 TAGGAAGAAACCCGTTACTTACCTACCGACATACAAATTTACCACTT 967

```

RESULT 10

BX370641

LOCUS

DEFINITION

BX370641 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED

ACCESSION

BX370641

VERSION

BX370641.1 GI:30459737

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens (human)

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 908)

AUTHORS

L.M.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7238.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0BA053ZH03 CS05016.1cluster=7238.f.

Contact: Peng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paraday Avenue Genoscope sequence ID: CS0BA053ZH03_CS05016.1.

Location/Qualifiers

1..908

/organism="Homo sapiens"

/mol_type="RNA"

/db_xref="taxon:9606"

/clone="CS0D004YP03"

/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"

/cell_line="RAMOS CELL LINE"

/clone_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT

25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 68.1%; Score 645.4; DB 13; Length 908;

Best Local Similarity 98.2%; Pred. No. 4.2e-136;

Matches 663; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

```

Qy 274 GAATGATGAACTTTTTCGAACCTGCAACTGATGATGATGATGATGATGATGATGATGATG 333
Db 1 GTATGCTGGAATTTTTCGAACCTGCAACTGATGATGATGATGATGATGATGATGATGATG 59
Qy 334 GATGCCCCCTTCTGCTGATATCAAGCGCTGCTGCTTTTATGAGGAATTTGATGATG 393
Db 60 GATGCCCCCTTCTGCTGATATCAAGCGCTGCTGCTTTTATGAGGAATTTGATGATG 119
Qy 394 CACTGAGAGGAAATGAGCAATAGTTCAAGTACCAATATATCAGGAACATGTTCAAC 453

```

Db 120 CACTGGAAGAAATGGACATTAGTTCAGAGTCAAGTATATCGAGAAACATGTTCAAC 179
Qy 454 CAATGCGCAAGTGGGTGTAACAGAGCAATGAAGACGATTTATATGACATGAT 513
Db 180 CAAATGCGCAAGTGGGTGTAACAGAGCAATGAAGACGATTTATATGACATGAT 239
Qy 514 GTAAAGCCAGCCGAGAAAGGGGCGAGAGACATGATTTGATTTCTACAGCTGTCCAA 573
Db 240 GTAAAGCCAGCCGAGAAAGGGGCGAGAGACATGATTTGATTTCTACAGCTGTCCAA 239
Qy 574 TTGTGTTAAGACCTTTAACAAGTTGCTGAATTTGAGACGAGATTCAGACATAGAA 633
Db 300 TTGTGTTAAGACCTTTAACAAGTTGCTGAATTTGAGACGAGATTCAGACATAGAA 339
Qy 634 ACCAATCTATACAGATATTTCTTTACAGTGAAGACCTATCTGGAATGAAGCA 693
Db 360 ACCAATCTATACAGATATTTCTTTACAGTGAAGACCTATCTGGAATGAAGCA 419
Qy 694 TCTGTTTGGGCGACAGAAACAGACTCTTGTTAGCCATTAAGATTTTATAC 753
Db 420 TCTGTTTGGGCGACAGAAACAGACTCTTGTTAGCCATTAAGATTTTATAC 479
Qy 754 CCCTTCACACCACTTTCGCACTAAGATTTCTGTTGAGTCTCTGCAATTTTAT 813
Db 480 CCCTTCACACCACTTTCGCACTAAGATTTCTGTTGAGTCTCTGCAATTTTAT 539
Qy 814 GCAGTATGTCACAAACAGTTCATTTGTTTATTTATTTGAATTTGTTTACCT 873
Db 540 GCAGTATGTCACAAACAGTTCATTTGTTTATTTATTTGAATTTGTTTACCT 599
Qy 874 ATGAATTCCTTTATTTAATAAATCATATGAAGAAATCCCTTACCTATCAGAAACAA 933
Db 600 ATGAATTCCTTTATTTAATAAATCATATGAAGAAATCCCTTACCTATCAGAAACAA 659
Qy 934 AACCTCTCTGTTT 948
Db 660 AACCTCTCTGTTT 674

RESULT 11
CA488543 924 bp mRNA linear EST 14-NOV-2002
LOCUS CA488543
DEFINITION AGSCOURT.10808884 MAPCL Homo sapiens cDNA clone IMAGE:6720309 5',
mRNA sequence.
ACCESSION CA488543.1 GI:24950702
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 924)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@db-remail.nih.gov
Tissue Procurement: Kristi A. Egland, Ira Pastan
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L14M14279 row: j column: 21
High quality sequence stop: 637.
Location/Qualifiers
1..924

FEATURES
Source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6720309"

/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-RME1, INCaP"
/lab_host="EMD10B"
/clone_lib="MAPCL"
/note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subcloned with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Egland, James J. Vincent, Robert Strausberg,
Bungkok Lee & Ira Pastan. Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

Query Match 67.8%; Score 643.2; DB 14; Length 924;
Best Local Similarity 96.6%; Pred. No. 1.3e-135;
Matches 657; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 1 CGGGTCTCGGCACTCCCTCCGCGCACTGCGCGGCTTCAAGCGCTTGACTTC 60
Db 106 CGGGTCTCGGCACTCCCTCCGCGCACTGCGCGGCTTCAAGCGCTTGACTTC 165
Qy 61 CGTCCAAACCTGATCTCTTATTTGCAAGCTATATCTTTCTGCAACTGGCTACCT 120
Db 166 CGTCCAAACCTGATCTCTTATTTGCAAGCTATATCTTTCTGCAACTGGCTACCT 225
Qy 121 ATCCAGTTATGAGGAGTGTATGATGATGATGATGATGATGATGATGATGATG 180
Db 226 ATCCAGTTATGAGGAGTGTATGATGATGATGATGATGATGATGATGATG 285
Qy 181 GAATTTAATATGAGACCTCTCGGACACTTGAAATATGATGATGATGATGATG 240
Db 286 GAATTTAATATGAGACCTCTCGGACACTTGAAATATGATGATGATGATGATG 345
Qy 241 AGAAGTACATTAATCTGCAAGAACTACACATGATGATGATGATGATGATG 300
Db 346 AGAAGTACATTAATCTGCAAGAACTACACATGATGATGATGATGATGATG 405
Qy 301 AACTGTACATTTCCCTCCGACCTCGAGCTGAATGATGATGATGATGATGATG 360
Db 406 AACTGTACATTTCCCTCCGACCTCGAGCTGAATGATGATGATGATGATGATG 465
Qy 361 GCTGCTGCTTTTGTGAGGAAATGATGATGATGATGATGATGATGATGATG 420
Db 466 GCTGCTGCTTTTGTGAGGAAATGATGATGATGATGATGATGATGATGATG 525
Qy 421 CAAGTACACATTAATCTGCAAGAACTACACATGATGATGATGATGATGATG 480
Db 526 CAAGTACACATTAATCTGCAAGAACTACACATGATGATGATGATGATGATG 585
Qy 481 AATGAAACAGAAATTAATTTGAGACATGATGATGATGATGATGATGATGATG 540
Db 586 AATGAAACAGAAATTAATTTGAGACATGATGATGATGATGATGATGATGATG 645
Qy 541 GAGACATGTTGATTTCTTACGACTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 600
Db 646 GAGACATGTTGATTTCTTACGACTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 705
Qy 601 GCGGAATTTGAGACAGGTTCAAGACATGATGATGATGATGATGATGATGATG 660
Db 706 GCGGAATTTGAGACAGGTTCAAGACATGATGATGATGATGATGATGATGATG 765
Qy 661 AGTGAAGAACTTATCTT 680
Db 766 AGTGAAGAACTTATCTT 785

RESULT 12
BI090566 869 bp mRNA linear EST 20-JUN-2001
LOCUS BI090566
DEFINITION 602855673F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996891 5',
mRNA sequence.
ACCESSION BI090566

```

ERSION      BI090566.1  GI:14508896
EXMORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE       NIH-MGC http://mgc.nci.nih.gov/
JOURNAL     National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT     Unpublished (1999)
            Contact: Robert Strauberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1023 row: a column: 20
            High quality sequence stop: 843.
            Location/Qualifiers
                1..869
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:498891"
                /cell_line="MGC36"
                /lab_host="DH10B"
                /note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 1.5 Kb. Library prepared by Life
                Technologies."
RIGIN
Query Match      66.4%; Score 629.6; DB 12; Length 869;
Best Local Similarity 96.8%; Pred. No. 1.6e-132;
Matches 706; Conservative 0; Mismatches 15; Indels 8; Gaps 5;
Y 1 CGGGTCTCGGGGATCCCTCCCGGCGCACTGGCCGCTCCCTACAGGCGCTTGGCTTC 60
b 142 CGGGTCTCGGGATCCCTCCCGGCGCACTGGCCGCTCCCTACAGGCGCTTGGCTTC 201
Y 61 CGTCCAAAACCTGATCTTATTTGTCAAGTAACTATCTTTCTGTCCAACTGGCTCACT 123
b 202 CGTCCAAAACCTGATCTTATTTGTCAAGTAACTATCTTTCTGTCCAACTGGCTCACT 261
Y 121 ATCCCGATATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 189
b 262 ATCCCGATATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 321
Y 181 GAATTTAAATATGAGACCTCTCTGGGACACTTGAATAATATGATGATGATGATGATG 240
b 322 GAATTTAAATATGAGACCTCTCTGGGACACTTGAATAATATGATGATGATGATGATG 381
Y 241 AGAAGTACATTAACTGGCAAGAACTACAAATGGAATGGAATGGAATGGAATGGAATG 300
b 382 AGAAGTACATTAACTGGCAAGAACTACAAATGGAATGGAATGGAATGGAATGGAATG 441
Y 301 AACTGTACATTTCCCATCTCCGACTGAAATGGAATGGAATGGAATGGAATGGAATG 360
b 442 AACTGTACATTTCCCATCTCCGACTGAAATGGAATGGAATGGAATGGAATGGAATG 501
Y 361 GCTGCTGCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 420
b 502 GCTGCTGCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 561
Y 421 CAAGTNGCACTATATGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 480
b 562 CAAGTNGCACTATATGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 620
Y 481 AATGAAACGAGATTTATATGAGACATGATGATGATGATGATGATGATGATGATGATG 538

```

```

Db 621 AATGAAACGAGATCTATTATGAGACATGAAATGTAAGGACCCAGAAACAGCGGC 680
Y 539 CAGAGACATGTTGATTTGATCTTACAGCTGTCCAAATTTGTTGTAAGACCTTTAAG 598
Db 681 ACAGACATGTTGATTTGATCTTACAGCTGTCCAAATTTGTTGTAAGACCTTTAAG 740
Y 599 TGGCTGAA-TTGGAGCAGAGTTCAAGAA- TGAACCACTATATACAGAAATATTTCT 656
Db 741 TGGCTGAACTGGAGCAGAGTTCAAGAACTATGAAACCAATATATCAAGATATTTCT 800
Y 657 TTAACATGG-AGAACCTACTTATCTGGGAAA-TGAACATCTGTTTGGGCGCAACAG 713
Db 801 TTAACATGGAGAAACCTACTTATCTGGGAAAATGAACCACTGTTTGGGCGCAACAG 860
Y 714 AAACAGAC 722
Db 861 AAACAGAC 869

RESULT 13
LOCUS      B1223533
DEFINITION B1223533 787 bp mRNA linear EST 11-JUN-2001
VERSION    B1223533
KEYWORDS   mRNA sequence.
SOURCE     B1223533.1 GI:14676977
ORGANISM   Homo sapiens (human)
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE      NIH-MGC http://mgc.nci.nih.gov/
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strauberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1253 row: n column: 06
            High quality sequence stop: 762.
            Location/Qualifiers
                1..787
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5105093"
                /tissue_type="cervical carcinoma cell line"
                /lab_host="DH10B"
                /note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 1.4 Kb. Library prepared by Life
                Technologies."
ORIGIN
Query Match      63.4%; Score 601; DB 12; Length 787;
Best Local Similarity 98.5%; Pred. No. 5.1e-126;
Matches 659; Conservative 0; Mismatches 5; Indels 5; Gaps 5;
Y 1 CGGGTCTCGGGGATCCCTCCCGGCGCACTGGCCGCTCCCTACAGGCGCTTGGCTTC 60
Db 116 CGGGTCTCGGGATCCCTCCCGGCGCACTGGCCGCTCCCTACAGGCGCTTGGCTTC 175
Y 61 CGTCCAAAACCTGATCTTATTTGTCAAGTAACTATCTTTCTGTCCAACTGGCTCACT 120
Db 176 CGTCCAAAACCTGATCTTATTTGTCAAGTAACTATCTTTCTGTCCAACTGGCTCACT 235

```

	Query Match	63.3%	Score 600;	DB 14;	Length 641;	
	Best Local Similarity	100.0%;	Pred. No. 8.8e-126;			
	Matches 600;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1 CGGGCTCGGGGCAATCCCTCCGGGCGGCACATCGGCGGGTGCCTCAGACGGCTTGACTTC	/mol_type="rRNA"				
Dd	42 CGGGTCTCGGGCATCCCTCCGGGCGGCACATCGGCGGGTGCCTCAGACGGCTTGACTTC	/db_xref="taxon:9606"				
QY	61 CGTCCAAACCTGATCTTAATTGTCAAAGCTAATATCTTTGTGCCACTGGCTCACT	/clone="IMAGE:5937013"				
Dd	102 CGTCCAAACCTGATCTTAATTGTCAAAGCTAATATCTTTGTGCCACTGGCTCACT	/sex="female"				
QY	121 ATCCCACTTAGAGAGGGGTGATGATGACATGAAGTTTTCGATTACAAGCCCCAGATATGG	/tissue_type="normal endometrium, late proliferative phase, cycle day 13"				
Dd	162 ATCCCACTTAGAGAGGGGTGATGATGACATGAAGTTTTCGATTACAAGCCCCAGATATGG	/lab_host="DH10B (rt-resistant)"				
QY	181 GAATTATAATAGAGA CCTCTCGGACACTTGAAAAATTATGATGATGCGCATTGATTC	/clone_1lb="NICHDS uc1"				
Dd	222 GAATTATAATAGAGAGCTCTCTCGGACACTTGAAAAATTATGATGATGCGCATTGATTC	/note="Organ: uterus; Vector: pCMV-SPORE.1 ccdB (Resgen, Invitrogen Corporation); Site_1: NotI; Site_2: EcoRV; Cloned unidirectionally from microquantly amounts of mRNA from normal endometrial tissue (late proliferative phase, cycle day 13). Average insert size 1.9 kb. Library constructed by Resgen (Invitrogen Corporation)."				
QY	241 AGAAGTACATTAACTGGCAAGAACTACACATGGAATGTAATGATCTTTCCAATTGGC					
Dd	282 AGAAGTACATTAACTGGCAAGAACTACACATGGAATGTAATGATCTTTCCAATTGGC					
QY	301 AACTGTACATTTCCCACTCCGACCCTGAATGGAATGCCCTTCTGTGTAATCAAGGC					
Dd	342 AACTGTACATTTCCCACTCCGACCCTGAATGGAATGCCCTTCTGTGTAATCAAGGC					
QY	361 GTCGCTGCTTTTTTGAAGGAATGATGATCTCACTGAAAGAAAATGGGACATTAGTT					
Dd	402 GTCGCTGCTTTTTTGAAGGAATGATGATCTCACTGAAAGAAAATGGGACATTAGTT					
QY	421 CAAGTACCACTATATACAGAAACATGTTTAAACCAATGGCAAGTGGTGAACAAGAAC					
Dd	462 CAAGTACCACTATATACAGAAACATGTTTAAACCAATGGCAAGTGGTGAACAAGAAC					
QY	481 AATGAAACAGAAATTTATATGAGACATGGAATGTAAAAACCAAGCCAGAAAAGGGGGCA					
Dd	522 AATGAAACAGAAATTTATATGAGACATGGAATGTAAAAACCAAGCCAGAAAAGGGGGCA					
QY	541 GAGACATGCTTTGATTCCTACGACTGTGCCAATTTGTGTTAAGAGACCTTTAACAAGTTG					
Dd	582 GAGACATGCTTTGATTCCTACGACTGTGCCAATTTGTGTTAAGAGACCTTTAACAAGTTG					

RESULT 15
BP982158

LOCUS BP982158 1098 bp mRNA linear EST 23-JAN-2001

DEFINITION 60230897.F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4400298 5', mRNA sequence.

ACCESSION BF982158

VERSION BF982158

KEYWORDS BR982158.1 GI:12384970 EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1098)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM10105 row: 9 column: 19
High quality sequence stop: 693.
Location/Qualifiers
1..1098

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:440298"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: Not1; Site_2: Sal1; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

Query Match 59.3%; Score 562.4; DB 10; Length 1098;
Best Local Similarity 88.2%; Pred. No. 2.9e-117;
Matches 660; Conservative 0; Mismatches 76; Indels 12; Gaps 4;

1 CGGGCTCGGGGATCCCTCCCGGGGCGCATCGCCGCTTCAAGCGCTTTGACTTC 60
2 CGGGCTCGGGGATCCCTCCCGGGGCGCATCGCCGCTTCAAGCGCTTTGACTTC 184
61 CGTCCAAACCTGATCCTTATTTGCAAGCTAGATATTCTTGTCCAACTGGCTCACT 120
185 CGTCCAAACCTGATCCTTATTTGCAAGCTAGATATTCTTGTCCAACTGGCTCACT 244
121 ATCCAGTTATGAGGGTATGATGATGATGATGATGATGATGATGATGATGATGAT 180
245 ATCCAGTTATGAGGGTATGATGATGATGATGATGATGATGATGATGATGATGAT 304
181 GAATTTAAATATGAGAGCTCCCTGGGACATTTGATGATGATGATGATGATGATGAT 240
305 GAATTTAAATATGAGAGCTCCCTGGGACATTTGATGATGATGATGATGATGATGAT 364
241 AGAAGTACATTAACTGCAAGAACTACACATGAAATGATGATGATGATGATGATGAT 300
365 AGAAGTACATTAACTGCAAGAACTACACATGAAATGATGATGATGATGATGATGAT 424
301 AACTGTACATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGAT 360
425 AACTGTACATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGAT 484
361 GTGGCCGCTTTTGAAGGGAATTGATGATGATGATGATGATGATGATGATGATGAT 420
485 GTGGCCGCTTTTGAAGGGAATTGATGATGATGATGATGATGATGATGATGATGAT 544
421 CAAGTAGCAACTATATCAGGAAACATGTCACCAATGCAAAAGTGGTGAACAGGAC 480
545 CAAGTAGCAACTATATCAGGAAACATGTCACCAATGCAAAAGTGGTGAACAGGAC 604
481 AATGAAACAGGAATTTATTTATGACATGGAATGTAAAGCCAGCCAGAAAAGGGGCA 540
605 AATGAAACAGGAATTTATTTCTGAGACTGGCTGTCAAGCCAG--CAAGAACAGGGGCA 663
541 GAGACATGGTTGATTCCTACGACTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 600

Db 664 GAGACCTGGTTGATTCCTACGACTGTTCCAAATT--GTGTACGAGACCTTAACAAGTTG 721
QY 601 GCTGAATTTGAGCAGAGTTCAAGAACATAGAACCACTATACAGAAATATTTCTTAC 660
Db 722 GCTGAATTTGAGCAGAGTTCAAGAACATAGAACCACTATACAGAAATATTTCTTAC 775
QY 661 AGTGAAGAACCTACTTATCTGGGAAATGAAACATCTGTTTGGGCCAACAGAAACAG 720
Db 776 GGGGAAACCTACTTCTTTGGCGGAGATGACACA---CTGGTGTTCGGGGCACCGGAAACCG 832
QY 721 ACTCTGTTTAGCCATTAAGAAATTTT 748
Db 833 CCCCTCGGTTCTCTCTCAAGAGAAATTT 860

Search completed: April 22, 2004, 04:39:48
Job time : 2858.76 secs

Db 661 CCCAGAAAAGGGGCGAGACATGGTTGATTCCTACGACTGTTCCAATTGTGTTAG 720
Qy 721 GACCTTTAAACAAGTTGGCTGAATTTGGGAGCAAGTCAAGAACATAGAAAACAATAC 780
Db 721 GACCTTTAAACAAGTTGGCTGAATTTGGGAGCAAGTCAAGAACATAGAAAACAATAC 780
Qy 781 AAGAAATATTTCTTACAGTGAAGAACCTTATCTGGAATGAAACATGCTGTTTGG 840
Db 781 AAGAAATATTTCTTACAGTGAAGAACCTTATCTGGAATGAAACATGCTGTTTGG 840
Qy 841 GCCAACAAGAAAACAAGACTTGGTTTAGCCATAAAAAAGATTTATTAACCCCTCAACC 900
Db 841 GCCAACAAGAAAACAAGACTTGGTTTAGCCATAAAAAAGATTTATTAACCCCTCAACC 900
Qy 901 ACATTTGCCAATTAAGAAATTTCTGTGAGTCTCTGCAAAATTTTGTGAGTGTGT 960
Db 901 ACATTTGCCAATTAAGAAATTTCTGTGAGTCTCTGCAAAATTTTGTGAGTGTGT 960
Qy 961 GCACAACAGTTCTATTGTTTATTAATTTGAATATGTTTATTAACCTATGAATCC 1020
Db 961 GCACAACAGTTCTATTGTTTATTAATTTGAATATGTTTATTAACCTATGAATCC 1020
Qy 1021 TTTTATTAATAATACATATGAGAATCCCTTTTACCTATCAGAAAACAACTCTCTGG 1080
Db 1021 TTTTATTAATAATACATATGAGAATCCCTTTTACCTATCAGAAAACAACTCTCTGG 1080
Qy 1081 TTTT 1084
Db 1081 TTTT 1084

RESULT 3
US-10-653-595-26
Sequence 26, Application US/10653595
Publication No. US20040048304A1
GENERAL INFORMATION:
APPLICANT: Ruben et. al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027P1C1
CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03
PRIOR APPLICATION NUMBER: US 09/397945
PRIOR FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/058604
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PAM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1520)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:

NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-10-653-595-26

Query Match 98.4%; Score 1067; DB 13; Length 1751;
Best Local Similarity 99.6%; Pred. No. 3 2e-278;
Matches 1067; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 14 GGCAGAGAGTAGACACGAGCAGGCGCCGAGATGGCGGAGCGCGGCGCTCGG 73
Db 16 GCGCAGAGAGTAGACACGAGCAGGCGCCGAGATGGCGGAGCGCGGCGCTCGG 75
Qy 74 GAGCGGCGCTTCCTGCTGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGG 133
Db 76 GAGCGGCGCTTCCTGCTGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGG 135
Qy 134 TCCCGGCTTCGCGGCACTCCCTCCCGGCGCACTGGCCGGTGCCTTACAAAGCCTTGGAC 193
Db 136 TCCCGGCTTCGCGGCACTCCCTCCCGGCGCACTGGCCGGTGCCTTACAAAGCCTTGGAC 195
Qy 194 TTCGCTCCAAACCTGATCCTTATTTGTCAAGCTAAGTATACCTTGTGCCACTGCTCA 253
Db 196 TTCGCTCCAAACCTGATCCTTATTTGTCAAGCTAAGTATACCTTGTGCCACTGCTCA 255
Qy 254 CCGATCCGATTAAGAGGGTGAATGATGACATGAGTTTGGATTTACAAAGCCCACTA 313
Db 256 CCGATCCGATTAAGAGGGTGAATGATGACATGAGTTTGGATTTACAAAGCCCACTA 315
Qy 314 TGGGAATTTAAATATGAGAACCTCTGGGACACTTGAATTTATGATGATGCAATTGGA 373
Db 316 TGGGAATTTAAATATGAGAACCTCTGGGACACTTGAATTTATGATGATGCAATTGGA 375
Qy 374 TTCAGAAATCAATTAATCTGGCAAGAACTACACATGGAATGTATGAACCTTTCCAACCT 433
Db 376 TTCAGAAATCAATTAATCTGGCAAGAACTACACATGGAATGTATGAACCTTTCCAACCT 435
Qy 434 GGCACCTGACATTTCCCATCTCCGACCTGAAATGAGAGCCCTTCTGTGATGCA 493
Db 436 GGCACCTGACATTTCCCATCTCCGACCTGAAATGAGAGCCCTTCTGTGATGCA 495
Qy 494 GCGGCTGCTGCTTTTGTGAGGAAATGATGATGTTCACTGGAAGGAAATGGGACATTA 553
Db 496 GCGGCTGCTGCTTTTGTGAGGAAATGATGATGTTCACTGGAAGGAAATGGGACATTA 555
Qy 554 GTTCAAGTACACATATATGAGAAACATGTTCAACCAATGGCAAAAGTGGTGAACAG 613
Db 556 GTTCAAGTACACATATATGAGAAACATGTTCAACCAATGGCAAAAGTGGTGAACAG 615
Qy 614 GACATGAAACAGGAATTTATATGAGACATGGAATGTAAAGCCAGCCCAAGAAAGGGG 673
Db 616 GACATGAAACAGGAATTTATATGAGACATGGAATGTAAAGCCAGCCCAAGAAAGGGG 675
Qy 674 GCAGAGACATGTTGATCTTACAGACTGTTCCAAATTTGTGTAAAGCACTTTTAAACAG 733
Db 676 GCAGAGACATGTTGATCTTACAGACTGTTCCAAATTTGTGTAAAGCACTTTTAAACAG 735

QY	734	TTGGCTGCAATTTGGAGCAGAGTTCAAGAAACATAGAAACCACTATACAGAAATATTCTT	793
Db	736	TTGGCTGAATTTGGAGCAGAGTTCAAGAAACATAGAAACCACTATACAGAAATATTCTT	795
QY	734	TACAGTGGGAACCTACCTATATCTGGGAAATGAAACATCTGTTTTGGGCAACAGAAAC	853
Db	736	TACAGTGGGAACCTACCTATATCTGGGAAATGAAACATCTGTTTTGGGCAACAGAAAC	855
QY	854	AAGACTCTGGTTTGGCCATAAAAAGATTTATTAACCCCTTCAACACACATTTGGCAACT	913
Db	856	AAGACTCTGGTTTGGCCATAAAAAGATTTATTAACCCCTTCAACACACATTTGGCAACT	915
QY	914	AAAGAAATTTCTGTAGAGTCTCTTGCAGAAATTTTGAATGACAGTATGTGCACAAACAGTTC	973
Db	916	AAAGAAATTTCTGTAGAGTCTCTTGCAGAAATTTTGAATGACAGTATGTGCACAAACAGTTC	975
QY	974	TATTTGTTTAAATTTTGAATATTTGGTTTAACTATGAATTTCCCTTTATTTAAATA	1033
Db	976	TATTTGTTTAAATTTTGAATATTTGGTTTAACTATGAATTTCCCTTTATTTAAATA	1033
QY	1034	ACATATGAAGAAATCCCTTAAACCTATCAAGAAACAAACACACCTCTGGTTTA	1084
Db	1036	ACATATGAAGAAATCCCTTAAACCTATCAAGAAACAAACACCTCTGGTTTA	1086
RESULT 4			
US-09-397-945-26			
; Sequence 26, Application US/09397945			
; Publication No. US20030065139A1			
GENERAL INFORMATION:			
; APPLICANT: Human Genome Sciences, Inc. et al.			
; TITLE OF INVENTION: 95 Human secreted proteins			
; FILE REFERENCE: P2027P1			
; CURRENT APPLICATION NUMBER: US/09/397,945			
; CURRENT FILING DATE: 1999-09-17			
; PRIOR APPLICATION NUMBER: PCT/US99/05804			
; PRIOR FILING DATE: 1999-03-18			
; PRIOR APPLICATION NUMBER: 60/078,566			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/078,576			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/078,573			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/078,574			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/078,579			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/080,314			
; PRIOR FILING DATE: 1998-04-01			
; PRIOR APPLICATION NUMBER: 60/080,312			
; PRIOR FILING DATE: 1998-04-01			
; PRIOR APPLICATION NUMBER: 60/078,578			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/078,581			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/078,577			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/078,563			
; PRIOR FILING DATE: 1998-03-19			
; PRIOR APPLICATION NUMBER: 60/080,313			
; PRIOR FILING DATE: 1998-04-01			
; NUMBER OF SEQ ID NOS: 470			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 26			
LENGTH: 1751			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURE:			
NAME/KEY: SITE			
LOCATION: (1520)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (1557)			

```

? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (1689)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (1729)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (1735)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (1741)
? OTHER INFORMATION: n equals a,t,g, or c
US-09-397-945-26

```

Query Match	98.4%	Score 1067;	DB 13;	Length 1751;
Best Local Similarity	99.6%	Pred. No. 3.2e-278;		
Matches 1067; Conservative	2;	Mismatches 2;	Indels 0;	Gaps 0

QY	14	GGGCAAGAGATGACACCGGCAACAGGGGCGCGAGATGCGCGGGGCGCGGGCGGCTCGG	73
Db	16	GGCAAGAGATGACACCGGCAACAGGGGCGCGAGATGCGGGGCGCGGGCGGCTCGG	75
QY	74	GGACGGCTTCCTGAGTGTGGGCGCTGGCGCTGCTTGGCTGGCGGTGGTTCGGGCTGG	133
Db	76	GGACGGCTTCCTGAGTGTGGGCGCTGGGCGCTTGGCTGGCGGTGGTTCGGGCTGG	135
QY	134	TCCGGGCTCGGGCATCCCTCCCGGGCGCATGGCGGGTCCCTACAGCGCTTGGAC	193
Db	136	TCCGGGCTCGGGCATCCCTCCCGGGCGCATGGCGGGTCCCTACAGCGCTTGGAC	195
QY	194	TTCCGCTCAAACCTGATTCCTATTGTCAAGCTAGATATCTTGTCCAACTGGCTCA	253
Db	196	TTCCGCTCAAACCTGATTCCTATTGTCAAGCTAGATATCTTGTCCAACTGGCTCA	255
QY	254	CCATATCCAGTATGAGGGGTGATATGACATGAAAGTTTTGCATTACAGGCCAGTA	313
Db	256	CCATATCCAGTATGAGGGGTGATATGACATGAAAGTTTTGCATTACAGGCCAGTA	315
QY	314	TGGGAATTTAAATATGAGACCTCCTGGGACACTTGAATAATATGCATATGCATTGA	373
Db	316	TGGGAATTTAAATATGAGACCTCCTGGGACACTTGAATAATATGCATATGCATTGA	375
QY	374	TTCAAGAGTACATTACTGGCAAGAACTACACAAATGGAATGGTATGAACTTTCCAACTT	433
Db	376	TTCAAGAGTACATTACTGGCAAGAACTACACAAATGGAATGGTATGAACTTTCCAACTT	435
QY	434	GGCACTGTACATTTCCCATCTCCGACTGAAATGATGCCCTTTCTGTGTATCA	493
Db	436	GGCACTGTACATTTCCCATCTCCGACTGAAATGATGCCCTTTCTGTGTATCA	495
QY	494	GGCGGTGCTGCTTTTGTGAGGGAATGATGATGTTCACTGGAGAAGAAATGGGACATTA	553
Db	496	GGCGGTGCTGCTTTTGTGAGGGAATGATGATGTTCACTGGAGAAGAAATGGGACATTA	555
QY	554	GTTCAAGTAGCAACTATATCAGAAACAATGTTCAACCAATGGCAAGTGGGTGAACAAG	613
Db	556	GTTCAAGTAGCAACTATATCAGAAACAATGTTCAACCAATGGCAAGTGGGTGAACAAG	615
QY	614	GACATGAAACAGGAATTTATTATGACATGGAATGTAAAGCCAGGCCAGAAAAAGGG	673
Db	616	GACATGAAACAGGAATTTATTATGACATGGAATGTAAAGCCAGGCCAGAAAAAGGG	675
QY	674	GAAGAGCATGATGTTGATTCACGACGTTCCAAATTTGGTAAAGACCTTTAACAG	733
Db	676	GAAGAGCATGATGTTGATTCACGACGTTCCAAATTTGGTAAAGACCTTTAACAG	735
QY	734	TTGGCTGAATTTGGAGCAGAGTTCCAGAACATAGAAACCACTATACAGAAATATTTCTT	793
Db	736	TTGGCTGAATTTGGAGCAGAGTTCCAGAACATAGAAACCACTATACAGAAATATTTCTT	795
QY	794	TACAGTAGAACCCTACTTATCTGGGAATGAACAATCTGTTTTGGGCGCAACAGAAAC	853

Db	796	TACAGTGAGAAACCTACTTATCTGGGAATGAACA	CATCTGTTTTGGGCAACA	GAGAAAC	855
Qy	854	AAAGCTCTGGTTTARGCATAAAAAGATTTATATTC	CCCCCTTCAACCA	CATTTGGCAACT	913
Db	856	AAGACTCTGGTTTARGCCATAAAAAGATTTATATTC	CCCCCTTCAACCA	CATTTGGCAACT	915
Qy	914	AAAGAAITTCCTGTGAGTCTCTGCAAAATTTT	TGATGAGAGATTTGTGACAAACAGATTC		973
Db	916	AAAGAAITTCCTGTGAGTCTCTGCAAAATTTT	TATGAGAGATTTGTGACAAACAGATTC		975
Qy	974	TATTTGTTTTAAATTTTGAATATTTGGTTTTT	ACCTATGAAATTCCTTTATTTAAATA		1033
Db	976	TATTTGTTTTAAATTTTGAATATTTGGTTTTT	ACCTATGAAATTCCTTTATTTAAATA		1035
Qy	1034	ACATATGAGAAATTCCTTTACCTATGAGAAACAAA	ACATCTCTGGTTTA		1084
Db	1036	ACATATGAGAAATTCCTTTACCTATGAGAAACAAA	ACATCTCTGGTTTA		1086

```

RESULT 5
US-09-122-383-13
; Sequence 13, Application US/09122383A
; Patent No. US20020042093A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; FILE REFERENCE: 97-38
; CURRENT APPLICATION NUMBER: US/09/122,383A
; EARLIER FILING DATE: 1998-07-24
; EARLIER APPLICATION NUMBER: 60/053,613
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate nucleotide sequence encoding zsig46
; OTHER INFORMATION: polypeptide of SEQ ID NO:2
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)..(1038)
; OTHER INFORMATION: n is any nucleotide
; US-09-122-383-13

```

Query Match	71.3%	Score 772.4;	DB 9;	Length 1038;
Best Local Similarity	61.7%	Pred. No. 1.7e-198;		
Matches 640;	Conservative 232;	Mismatches 165;	Indels 0;	Gaps 0;

QY 4 ATCCGCGGGAGCGCGGGCGCGCGCTCTGGGGAGAGCGCTTCCTGGTCTCTGGGCGCTGGCGCTG 106

Db 1 ATCGMGMGMGNCNGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 60

QY 107 CTTTGGCTCGCGGTGGTTCCGGGCTGGTCCGGGCTCTCGGGCATCCCTCCCGCGGCAC 166

Db 61 YTTGTGYTNGCGTNGTNGTNGCGNNGTNGTNGSNNNGTNGSNGNATHTCNMSMGMGMCA 120

QY 167 TGGCGGGTCCCTACAGAGGCTTTGACTTCGGTCCAAACCTGATCCTCTGTGTGTCAAGCT 226

Db 121 TGGCGGNTGNCNTAATAARMGNTTGAATTTGNGNCNNAACCNAGACNTATATGCAAGCA 180

QY 227 AAGTATACCTTTCTGTCCAACTGGCTCACTTATCCAGTTATGAGAGGTGATGATGACATT 286

Db 181 AATAATACNTTTTGGCCNAACNGMNSNCNATHTCCNGTATGAGAGGNGAGYGAATHT 240

QY 287 GAAGTTTTGCATTCAAGCCCAAGTATGGGAATTTAAATTTGAGAACCTCCAGGAGCAC 346

QY 241 GAGTNTTATGNTTNCARGCNCNCNGTNGGGAATTTAATATGAGNGAYTNTTNGNCAT 300

QY 347 TTGAATAATTATGATGATGCCATTGGATTCAAGAGTCAATTAACCTGGCAAGAACTACACA 406

Db	301	YTNAAATATGAAGCAGAACTGGNATTYMGNSNACNTYTNACNCGGAATAAAATATAACN	360
QY	407	ATGGAATGGTATGAACTTTCCAACTTGGCACTGTACATTCCCCCATCTCCGACTGAA	466
Db	361	ATGGAATGGTATGAACTTTCCAACTTGGCACTGTACATTCCCCCATCTCCGACTGAA	420
QY	467	ATGATGCCCCCTTCTGGTGTATCATCAAGAGCGCTCCGCTCTTTTGGAGGAATATGATAT	526
Db	421	ATGATGAGCNCNTTYYTGTTGTAACAAGAGCNCNCNTGYTTTYYTGAGGNAITHAYATY	480
QY	527	GTTCACTGGAAGAAATGGCACTTAGTTCACAGTAGCACTATATCAGAAACATGTTTC	586
Db	481	GTTCACTGGAAGAAATGGCACTTAGTTCACAGTAGCACTATATCAGAAACATGTTTC	540
QY	587	AACCAANTGGCAAGTGGTGAAACAGGACAACTGAAACAGGAATTTATATGACATGG	646
Db	541	AAACCAANTGGCAAGTGGTGAAACAGGACAACTGAAACAGGAATTTATATGACATGG	600
QY	647	AATGTAAAGCCAGCCAGAAAGGGGGCAGAGACATGTTTGATTCCTACGACGTGTC	706
Db	601	AAATGTAAAGCCAGCCAGAAAGGGGGCAGAGACATGTTTGATTCCTACGACGTGTC	660
QY	707	AAATTTGTGTTAAGAACCTTTAAACAAGTGGCTGAATTTGGACAGAGTTCAAGACATA	766
Db	661	AAATTTGTGTTAAGAACCTTTAAACAAGTGGCTGAATTTGGACAGAGTTCAAGACATA	720
QY	767	GAAACCAACTTACAAAGAATTTCTTTACAGTGGAGAACCTTACTTATCTGGGAAATGAA	826
Db	721	GAAACCAACTTACAAAGAATTTCTTTACAGTGGAGAACCTTACTTATCTGGGAAATGAA	780
QY	827	ACATCTGTTTTGGGCCAACAAGAAACAAGCTCTTGTTAGCCATTAATAAAGATTTTAT	886
Db	781	ACATCTGTTTTGGGCCAACAAGAAACAAGCTCTTGTTAGCCATTAATAAAGATTTTAT	840
QY	887	TACCCCTTCAAAACCACTTTGCCAACAATAAAGATTTCTTGAGTCTCTTGCAAAATTTT	946
Db	841	TACCCCTTCAAAACCACTTTGCCAACAATAAAGATTTCTTGAGTCTCTTGCAAAATTTT	900
QY	947	GATCAGCTGATGTGCACAACAGCTTCTATTGTTTATATTTGAATATGTTGTTTGA	1006
Db	901	GATCAGCTGATGTGCACAACAGCTTCTATTGTTTATATTTGAATATGTTGTTTGA	960
QY	1007	CTATGAAATTCCTTTTATTTAAATATACATATGAGAGAAATCCCTTACTATACAGAAC	1066
Db	961	CTATGAAATTCCTTTTATTTAAATATATACATATGAGAGAAATCCCTTACTATACAGAAC	1020
QY	1067	AAAACACTCTCTGTTT 1083	
Db	1021	AAAACACTCTCTGTTT 1037	

RESULT 6
US-10-010-050A-13
Sequence 13, Application US/10010050A
Publication NO. US20020173624A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
TITLE OR INVENTION: CHROMOSOME 13
FILE REFERENCE: 97-38CI
CURRENT APPLICATION NUMBER: US/10/010,050A
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 09/122,383
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 60/053,613
PRIOR FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 1038
TYPE: DNA

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: NT HIT: 915729771, EVALUATE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: 075503, EVALUATE 2.00e-96
OTHER INFORMATION: EST_HUMAN HIT: H01255.1, EVALUATE 0.00e+00
US-09-864-761-22592

Query Match 45.3%; Score 491; DB 9; Length 494;
Best Local Similarity 100.0%; Pred. No. 2e-122;
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 594 TGGCAAAAGTGGTGAAGCAAGCAATGAAGCAATTTATGAGACATGAAATGTA 653
DB 1 TGGCAAAAGTGGTGAAGCAAGCAATGAAGCAATTTATGAGACATGAAATGTA 60
QY 654 AACCCAGCCAGAAAAGGGGAGAGACATGTTGATCTGAGCTGTTCCAAATTG 713
DB 61 AACCCAGCCAGAAAAGGGGAGAGACATGTTGATCTGAGCTGTTCCAAATTG 120
QY 714 TGTAAAGACCTTTAAACAAGTGGTGAATTTGAGAGAGTTCAAGAACATGAACA 773
DB 121 TGTAAAGACCTTTAAACAAGTGGTGAATTTGAGAGAGTTCAAGAACATGAACA 180
QY 774 ACTATCAAGAAATTTCTTTAAGTGAAGCACTTATCTGAGGAATGAACAATCTG 833
DB 181 ACTATCAAGAAATTTCTTTAAGTGAAGCACTTATCTGAGGAATGAACAATCTG 240
QY 834 TTTTGGGCGAACAAGAAACAAGCTTGGTTAGCAATTAAGATTTATACCCCT 693
DB 241 TTTTGGGCGAACAAGAAACAAGCTTGGTTAGCAATTAAGATTTATACCCCT 300
QY 894 TCAACCAATTTGCCAATTAAGAAATTTCTGTTGAGTCTCTGCAATTTTGTATGAC 553
DB 301 TCAACCAATTTGCCAATTAAGAAATTTCTGTTGAGTCTCTGCAATTTTGTATGAC 360
QY 954 TGAATGTGACAAACAGTCTATTTGTTTATATTTGATTTGATTTGATTTGATTA 1013
DB 361 TGAATGTGACAAACAGTCTATTTGTTTATATTTGATTTGATTTGATTTGATTA 420
QY 1014 AATCCCTTTTATTAATATACATATGAGAAATCCCTTACCTATCAAGAAACAAAC 1073
DB 421 AATCCCTTTTATTAATATACATATGAGAAATCCCTTACCTATCAAGAAACAAAC 480
QY 1074 TCTCTGGTTTA 1084
DB 481 TCTCTGGTTTA 491

RESULT 8

US-10-040-739-825
Sequence 825, Application US/10040739
Publication No. US2002017363A1
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
McCoy, John
LaValle, Edward
Racie, Lisa
Merberg, David
Treacy, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
NUMBER OF SEQUENCES: 1519
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.

ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040.739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 825:
SEQUENCE CHARACTERISTICS:
LENGTH: 506 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 825:
US-10-040-739-825

Query Match 37.0%; Score 401.4; DB 14; Length 506;
Best Local Similarity 99.8%; Pred. No. 4e-98;
Matches 402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 497 GCTGCTGCTTTTGGAGGAATGATGATGTTCACTGAGAGAAATGAGCAATTAGTT 556
DB 99 GCTGCTGCTTTTGGAGGAATGATGATGTTCACTGAGAGAAATGAGCAATTAGTT 158
QY 557 CAAGTAGCACTATATCAAGAAACATGTTCAACCAATGCGAAAGTGGTGAACAGGAC 616
DB 159 CAAGTAGCACTATATCAAGAAACATGTTCAACCAATGCGAAAGTGGTGAACAGGAC 218
QY 617 AATGAAACAGAAATTAATATGAGCATGAGATGTAAGCCAGCCCAAGAAAGGGGCA 676
DB 219 AATGAAACAGAAATTAATATGAGCATGAGATGTAAGCCAGCCCAAGAAAGGGGCA 278
QY 677 GAGACATGGTTGATTCCTACGACTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 736
DB 279 GAGACATGGTTGATTCCTACGACTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 338
QY 737 GCTGAATTTGAGCGAGGTTCAAGAACATGAACCAACTATACAAAGATTTCTTAC 796
DB 339 GCTGAATTTGAGCGAGGTTCAAGAACATGAACCAACTATACAAAGATTTCTTAC 398
QY 797 AGTGAGAACCTATCTATCTGAGAAATGAAACATGTTTTTGGGCGAACAAGAAACAG 856
DB 399 AGTGAGAACCTATCTATCTGAGAAATGAAACATGTTTTTGGGCGAACAAGAAACAG 458
QY 857 ACTCTGTTTGAACCAATTAAGATTTTATTTACCCCTTCAAC 899
DB 459 ACTCTGTTTGAACCAATTAAGATTTTATTTACCCCTTCAAC 501

RESULT 9

US-09-864-761-5832
Sequence 5832, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```
/ FILE REFERENCE: Aecmca-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 5832
/ LENGTH: 474
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC001226.1
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
/ OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
/ US-09-864-761-5832

Query Match      35.1%; Score 380.2; DB 9; Length 474;
Best Local Similarity 99.2%; Pred No. 2,1e-92;
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 568 TATATCAGAAACATGTTCAACCAATGCAAAAGTGCGTGAACACAGACCAATGAAACAGG 627
DB 90 TAACTAGAGAAACATGTTCAACCAATGCAAAAGTGCGTGAACACAGGCAATGAAACAGG 149
QY 628 AATTATATGACATGGAATGTAATAAGCCCAAGAAAGGGGGGCGAGACATGTT 687
DB 150 AATTATATGACATGGAATGTAATAAGCCCAAGAAAGGGGGGCGAGACATGTT 209
QY 688 TGATTCCTGACACTGTTCCAAATTTGTGTTAAGACCTTTAAACAAGTGGCTGAATTTGG 747
DB 210 TGATTCCTGACACTGTTCCAAATTTGTGTTAAGACCTTTAAACAAGTGGCTGAATTTGG 269
```

```
QY 748 AGCAGAGTTCAAGACATGAGAACCAACTATACAAAGATATTTCTTTACAGTGGAGAAC 807
DB 270 AGCAGAGTTCAAGACATGAGAACCAACTATACAAAGATATTTCTTTACAGTGGAGAAC 329
QY 808 TACTTATATGGAATGAAATGTAATCTGTTTGGGCGCAACAGAAACAGACTCTGGTT 867
DB 330 TACTTATATGGAATGAAATGTAATCTGTTTGGGCGCAACAGAAACAGACTCTGGTT 389
QY 868 AGCCATTAAGATTTTATATACCCCTTCAACCAACATTTGCCAATAAGATTTCTGTT 927
DB 390 AGCCATTAAGATTTTATATACCCCTTCAACCAACATTTGCCAATAAGATTTCTGTT 449
QY 928 GAGTCTCTGCAAAATTTTGGATGCA 952
DB 450 GAGTCTCTGCAAAATTTTGGATGCA 474

RESULT 10
US-10-264-049-1859
Sequence 1859, Application US/10264049
Publication No. US20040005579A1
GENERAL INFORMATION:
APPLICANT: Biree et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PAL33P1
CURRENT APPLICATION NUMBER: US/10/264,049
CURRENT FILING DATE: 2002-10-04
PRIOR APPLICATION NUMBER: PCT/US01/18569
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: US 60/209,467
PRIOR FILING DATE: 2000-06-07
NUMBER OF SEQ ID NOS: 4360
SOFTWARE: Patent In Ver. 3.1
SEQ ID NO 1859
LENGTH: 497
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (175)..(175)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (175)..(175)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (285)..(285)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (295)..(295)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (313)..(313)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (323)..(324)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (331)..(331)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (395)..(395)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (440)..(440)
OTHER INFORMATION: n equals a,t,g, or c
```

FEATURE:
NAME/KEY: misc feature
LOCATION: (458)..(458)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (464)..(464)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (478)..(478)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (483)..(483)
OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-1859

Query Match 24.5%; Score 266; DB 16; Length 497;
Best Local Similarity 99.6%; Pred. No. 2e-61; Indels 0; Gaps 0;
Matches 266; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 818 GGAATGAAACATCTGTTTGGGCCAAGAAAGAGCTCTGTTAGCCATPAAA 877
DB 3 GGAATGAAACATCTGTTTGGGCCAAGAAAGAGCTCTGTTAGCCATPAAA 62
QY 878 AGATTTATTAACCCCTTCAACACATTTGGCACTPAAAGATTTCTGAGTCTTG 937
DB 63 AGATTTATTAACCCCTTCAACACATTTGGCACTPAAAGATTTCTGAGTCTTG 122
QY 938 CAATTTTATGATGAGATTTGCAACAAAGCTTCTATTTGTTATATTTTGAATAT 997
DB 123 CAATTTTATGATGAGATTTGCAACAAAGCTTCTATTTGTTATATTTTGAATAT 182
QY 998 TGGTTTATTAACCTTGAATTCCTTTTATTAATAATATGAGAAATCCCTTACCT 1057
DB 183 TGGTTTATTAACCTTGAATTCCTTTTATTAATAATATGAGAAATCCCTTACCT 242
QY 1058 ATCAGAAACAAACACTCTCTGTTTA 1084
DB 243 ATCAGAAACAAACACTCTCTGTTTA 269

RESULT 11
US-09-867-701-2579
Sequence 2579, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
APPLICANT: Agiate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2579
LENGTH: 351
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-2579

Query Match 18.2%; Score 197.2; DB 9; Length 351;
Best Local Similarity 84.2%; Pred. No. 7.2e-43;
Matches 234; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 378 GAAGTACATTAACTGGGAAAGAACTACACATGGAATGATGAACTTTTCCAACTGGCA 437
DB 1 GAAGTACATTAACTGGGAAAGAACTACACATGGAATGATGAACTTTTCCAACTGGCA 60
QY 438 ACTGACATTCCCACTCTCCGACCTGAATGATGATGCCCTTCTGTTGTAATCAAGGCG 497

DB 61 ACTGACATTCCCACTCTCCGACCTGAATGATGATGCCCTTCTGTTGTAATCAAGGCG 120
QY 498 CTGCTGCTTTTGGAGGAATTTGATGATGTTCTACTGGAAGAAATGGACATTAGTTC 557
DB 121 CTG-CTGCTTTTGGAGGAATTTGATGATGTTCTACTGGAAGAAATGGACATTAGTTC 179
QY 558 AAGTAGCACTATACAGAAACATGTTCAACCAATGGCAAGTGGTGAACAGAGCA 617
DB 180 AAGTAGCACTATACAGAAATGAGTGAATATGCAATATTTGATCATTCATCAAAA 239
QY 618 ATGAACAGAAATTTATATGACATGGAATGTAAA 655
DB 240 AACCAATGAAGAAATTTATATCTTCATTTGAAACA 277

RESULT 12
US-09-954-456-1363/c
Sequence 1363, Application US/09954456
Patent No. US20020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cance
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
PRIOR FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1363
LENGTH: 473
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-1363

Query Match 12.3%; Score 133; DB 9; Length 473;
Best Local Similarity 100.0%; Pred. No. 2.3e-25;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 952 AGGATTTGCAACAAAGCTTCTATTTGTTATATTTGAATATGTTGTTTACCTAT 1011
DB 473 AGGATTTGCAACAAAGCTTCTATTTGTTATATTTGAATATGTTGTTTACCTAT 414
QY 1012 GAAATCCCTTTTATTAATAATACATATGAGAAATCCCTTACCTATGAGAAACAAAC 1071
DB 413 GAAATCCCTTTTATTAATAATACATATGAGAAATCCCTTACCTATGAGAAACAAAC 354
QY 1072 ACTCTGCTGTTTA 1084
DB 353 ACTCTGCTGTTTA 341

RESULT 13

US-09-908-975-10339
; Sequence 10339, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10339
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-10339

Query Match 5.5%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.4e-06;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 438 ACTGATCTTCCCATCTCCGACCTGAAATGATGCCCCCTTGTGTATATCAAGCG 497
Db 1 ACTGATCTTCCCATCTCCGACCTGAAATGATGCCCCCTTGTGTATATCAAGCG 60
RESULT 14
US-09-938-842A-5232
; Sequence 5232, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5232
; LENGTH: 1945
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-5232

Query Match 4.2%; Score 45.4; DB 9; Length 1945;
Best Local Similarity 49.0%; Pred. No. 0.33;
Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 760 GAACATGAACCAACTATACAGATATTTCTTTACAGTGAGAACTACTATTCGGG 819
Db 121 GAATATGATTTCACTCCCTGATATATGTTATATGAAATGAGTTTACAAAGGTG 180
Qy 820 AAATGAACATCTGTTTGGGCCAAGAAAGAGCTGTGTTAGCCATAAAG 879
Db 181 AAATGCTTTTCATGTGCTTACCAAAAAAATGCTCTTTCATGATGTGCTTAAAGT 240

Qy 880 ATTTATACCCCTTCAACACATTTGCCACATAAAGATTTCTGTGAGTCTTGCA 939
Db 241 GTTATATGACCTTACAGCAACATGTAGAGATTTGAAATGACATGAGTCTTGAG 300
Qy 940 AATTTTGATGAGATTTGCAACAACAGTTCTATTTGTTTAAATTTGAATTTG 999
Db 301 AATGTGAAGCTGTGTGCTTGACCAAGAAGTTCTACTATTTGACTATTTGATAG 360
Qy 1000 GTTTTAA 1006
Db 361 GCCTGTA 367

RESULT 15
US-09-938-842A-5232
; Sequence 5232, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5232
; LENGTH: 1945
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-5232

Query Match 4.2%; Score 45.4; DB 11; Length 1945;
Best Local Similarity 49.0%; Pred. No. 0.33;
Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 760 GAACATGAACCAACTATACAGATATTTCTTTACAGTGAGAACTACTATTCGGG 819
Db 121 GAATATGATTTCACTCCCTGATATATGTTATATGAAATGAGTTTACAAAGGTG 180
Qy 820 AAATGAACATCTGTTTGGGCCAAGAAAGAGCTGTGTTAGCCATAAAG 879
Db 181 AAATGCTTTTCATGTGCTTACCAAAAAAATGCTCTTTCATGATGTGCTTAAAGT 240
Qy 880 ATTTATACCCCTTCAACACATTTGCCACATAAAGATTTCTGTGAGTCTTGCA 939
Db 241 GTTATATGACCTTACAGCAACATGTAGAGATTTGAAATGACATGAGTCTTGAG 300
Qy 940 AATTTTGATGAGATTTGCAACAACAGTTCTATTTGTTTAAATTTGAATTTG 999
Db 301 AATGTGAAGCTGTGTGCTTGACCAAGAAGTTCTACTATTTGACTATTTGATAG 360
Qy 1000 GTTTTAA 1006
Db 361 GCCTGTA 367

Search completed: April 22, 2004, 05:09:05
Job time : 848.604 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 00:17:57 ; Search time 92.2894 Seconds

(without alignments)
6518.267 Million cell updates/sec

Title: US-10-010-050A-1_COPY_1_1084

Perfect score: 1084
Sequence: 1 gaattcgccagcagggcagcag.....acaacacactctctgttca 1084

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	50.6	4.7	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	41.2	3.8	640681	4 US-09-790-988-1	Sequence 1, Appl
C 3	40.6	3.7	2722	2 US-08-500-857A-7	Sequence 7, Appl
C 4	40	3.7	3877	2 US-08-599-895-1	Sequence 1, Appl
C 5	40	3.7	3877	3 US-09-211-290-1	Sequence 1, Appl
C 6	40	3.7	3877	3 US-09-322-676-1	Sequence 1, Appl
C 7	40	3.7	3877	3 US-09-466-036A-1	Sequence 1, Appl
C 8	39.6	3.7	2428	2 US-08-849-536A-3	Sequence 3, Appl
C 9	39.2	3.6	832	4 US-09-621-976-2813	Sequence 2813, Ap
C 10	39.2	3.6	8607	4 US-10-204-708-41	Sequence 41, Appl
C 11	39.2	3.6	21721	4 US-09-269-939A-41	Sequence 19, Appl
C 12	39.2	3.6	22976	4 US-09-269-939A-19	Sequence 1, Appl
C 13	39.2	3.6	23187	4 US-09-459-522-1	Sequence 1, Appl
C 14	39	3.6	429	4 US-09-252-991A-10838	Sequence 10838, A
C 15	39	3.6	474	4 US-09-252-991A-10890	Sequence 18033, A
C 16	38.2	3.5	786431	4 US-09-751-389-3	Sequence 3, Appl
C 17	37.8	3.5	543	4 US-08-956-171E-907	Sequence 907, App
C 18	37.6	3.5	4257	2 US-08-680-473-1	Sequence 1, Appl
C 19	37.6	3.5	4257	3 US-09-259-821A-1	Sequence 1, Appl
C 20	37.6	3.5	4257	3 US-08-843-659-1	Sequence 1, Appl
C 21	37.6	3.5	12001	1 US-08-458-568A-11	Sequence 11, Appl
C 22	37.6	3.5	4403765	3 US-09-103-840A-2	Sequence 2, Appl
C 23	37.2	3.4	4411529	3 US-08-675-773B-4	Sequence 1, Appl
C 24	37.2	3.4	2268	4 US-08-545-528D-1	Sequence 4, Appl
C 25	36.8	3.4	580073	4 US-09-103-840A-2	Sequence 1, Appl
C 26	36.8	3.4	4403765	3 US-09-103-840A-2	Sequence 2, Appl
C 27	36.8	3.4	4403765	3 US-09-103-840A-2	Sequence 2, Appl

C 28	36.8	3.4	4411529	3 US-09-103-840A-1	Sequence 1, Appl
C 29	36.6	3.4	1830121	4 US-09-557-884-1	Sequence 1, Appl
C 30	36.6	3.4	1830121	4 US-09-643-990A-1	Sequence 1, Appl
C 31	36.4	3.4	1767	4 US-09-016-434-1468	Sequence 1, Appl
C 32	36.4	3.4	1664976	4 US-08-916-421B-1	Sequence 1, Appl
C 33	36.2	3.3	9347	4 US-10-204-708-36	Sequence 36, Appl
C 34	36	3.3	168	3 US-08-897-924A-1	Sequence 1, Appl
C 35	36	3.3	364	4 US-09-621-976-17202	Sequence 17202, A
C 36	36	3.3	6757	4 US-09-023-655-1087	Sequence 1087, Ap
C 37	36	3.3	13086	4 US-08-956-171E-16	Sequence 16, Appl
C 38	36	3.3	43280	2 US-08-804-227C-1	Sequence 1, Appl
C 39	35.6	3.3	363	4 US-09-601-198-169	Sequence 169, App
C 40	35.4	3.3	447	4 US-09-319-014B-1	Sequence 1, Appl
C 41	35.4	3.3	832	4 US-09-621-976-2813	Sequence 2813, Ap
C 42	35.4	3.3	1164	1 US-08-416-478A-5	Sequence 5, Appl
C 43	35.4	3.3	1164	2 US-08-474-988B-5	Sequence 5, Appl
C 44	35.4	3.3	1164	2 US-08-394-442B-5	Sequence 5, Appl
C 45	35.4	3.3	1654	4 US-09-620-312D-581	Sequence 581, App

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 4.7%; Score 50.6; DB 1; Length 7218;

[illegible][illegible]

Db 698 AGCTGTTGATTACAGAAAGAAATTCAAGATTCTCCAGAACAGTGTGAAGTCTGT 757
QY 826 AACATCTGTTTGGGCCCAACAGAAACAGACTTGTGTTAGCCA 872
Db 758 TTCAGTTGTTCTATGTTATGTTAATATGAGGTAATGATTGATTCAGCTA 804

RESULT 4
US-08-599-895-1
; Sequence 1, Application US/08599895
; Patent No. 5891855
; GENERAL INFORMATION:
; APPLICANT: Florkiewicz, Robert Z.
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,895
; FILING DATE: 31-JAN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5891855endburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 760100.416
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3877 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-599-895-1

Query Match 3.7%; Score 40; DB 2; Length 3877;
Best Local Similarity 55.9%; Pred. No. 0.37;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 54 GGGGCGGGGCGGCGCTCGGGGAGCGGCTTCTGTGCTGGGCGGCTGCTTTGGC 113
Db 221 GGGGGGCGGGGCGGGGCGGCTGCCCCGAGAGCGGCTCGAGGCGCGGCGCGGGG 280

QY 114 TCGCGGTGTTCCGGGCTGTGCTCCGGGCTCTCGGGCATCCCTCCGCGGCACTGGCCGG 173
Db 281 ACGGCGGCTCCCGCGGCTCCAGCGGCTTCGGGATCCCGCGGCGCGCGAGACCA 340

QY 174 TGGCCTTACAAGCGCTT 189
Db 341 TGGCAGCCGGGAGCAT 356

RESULT 5
US-09-211-290-1
; Sequence 1, Application US/09211290
; Patent No. 6071885
; GENERAL INFORMATION:
; APPLICANT: Florkiewicz, Robert Z.
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/211,290
; FILING DATE: 12-DEC-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 200124.401D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3877 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-211-290-1

Query Match 3.7%; Score 40; DB 3; Length 3877;
Best Local Similarity 55.9%; Pred. No. 0.37;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 54 GGGGCGGGGCGGCGCTCGGGGAGCGGCTTCTGTGCTGGGCGGCTGCGGCTTTGGC 113
Db 221 GGGGGGCGGGGCGGGGCGGCTGCCCCGAGAGCGGCTCGAGGCGCGGCGCGGGG 280

QY 114 TCGCGGTGTTCCGGGCTGTGCTCCGGGCTCTCGGGCATCCCTCCGCGGCACTGGCCGG 173
Db 281 ACGGCGGCTCCCGCGGCTCCAGCGGCTTCGGGATCCCGCGGCGCGCGAGACCA 340

QY 174 TGGCCTTACAAGCGCTT 189
Db 341 TGGCAGCCGGGAGCAT 356

RESULT 6
US-09-322-676-1
; Sequence 1, Application US/09322676
; Patent No. 6107283
; GENERAL INFORMATION:
; APPLICANT: Florkiewicz, Robert Z.
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/322,676
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/211,290
; FILING DATE: 12-DEC-1998

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 200124.401D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3877 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-322-676-1

Query Match 3.7%; Score 40; DB 3; Length 3877;
Best Local Similarity 55.9%; Pred. No. 0.37;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 54 GGGGCGGGGCGGCGCTCGGGAGCGCGCTTCGTGCTGCGCGCTGCGCTTGGC 113
DB 221 GGGGGGCGGGGCGGGGCGGCGCTGCCCCGAGCGGCTCGAGGCGGGGCGGGGG 280
QY 114 TCGGGGTGTTCCGGGCTGCTCCCGGCTTCGCGGCACTCCCTCCCGCGCACTGGCCGG 173
DB 281 ACGGCGGCTCCCGCGGCTTCAGCGGCTTCGGGATCCCGCGCGCGCGCGAGCA 340
QY 174 TGCCCTACAGCGCTT 189
DB 341 TGGCAGCGGGAGCAT 356

RESULT 7

US-09-466-036A-1

; Sequence 1, Application US/09466036A

; Patent No. 6281197

; GENERAL INFORMATION:

APPLICANT: Flokiewicz, Robert Z.
TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/466,036A

FILING DATE: 17-Dec-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/211,290

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 200124.401D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3877 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-466-036A-1

Query Match 3.7%; Score 40; DB 3; Length 3877;
Best Local Similarity 55.9%; Pred. No. 0.37;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 54 GGGGCGGGGCGGCGCTCGGGAGCGCGCTTCGTGCTGCGCGCTGCGCTTGGC 113
DB 221 GGGGGGCGGGGCGGGGCGGCGCTGCCCCGAGCGGCTCGAGGCGGGGCGGGGG 280
QY 114 TCGGGGTGTTCCGGGCTGCTCCCGGCTTCGCGGCACTCCCTCCCGCGCACTGGCCGG 173
DB 281 ACGGCGGCTCCCGCGGCTTCAGCGGCTTCGGGATCCCGCGCGCGCGAGCA 340
QY 174 TGCCCTACAGCGCTT 189
DB 341 TGGCAGCGGGAGCAT 356

RESULT 8

US-08-849-536A-3

; Sequence 3, Application US/08849536A

; Patent No. 5853976

; GENERAL INFORMATION:

APPLICANT: HESSE, Friederike

APPLICANT: AMBROSIOUS, Dorothee

APPLICANT: BURTSCHER, Helmut

TITLE OF INVENTION: RECOMBINANT PROTEINASE FROM CLOSTRIDIUM

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP

STREET: 655 15th St., N.W., Suite 330 - G St. Lobby

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-5701

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/849,536A

FILING DATE: Herewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mong, King L.

REGISTRATION NUMBER: 37,500

REFERENCE/DOCKET NUMBER: 1614-7026

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 638 - 5000

TELEFAX: (202) 638 - 4810

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2428 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: sig_peptide

LOCATION: 970..1026

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 1027..1965

US-08-849-536A-3

Query Match 3.7%; Score 39.6; DB 2; Length 2428;
Best Local Similarity 51.7%; Pred. No. 0.37; 84; Indels 0; Gaps 0;
Matches 90; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 879 GATTATATACCCCTTCAACACCATTTGCCAATAAGATTTCTGTGAGTCTTGGC 938


```

NAME/KEY: exon
LOCATION: 12064..12183
NAME/KEY: exon
LOCATION: 15049..15105
NAME/KEY: exon
LOCATION: 15670..15816
NAME/KEY: exon
LOCATION: 19486..19659
NAME/KEY: exon
LOCATION: 19806..19865
NAME/KEY: exon
LOCATION: 19963..20094
NAME/KEY: exon
LOCATION: 20236..20864
NAME/KEY: exon
LOCATION: 20954..21094
NAME/KEY: Misc_Feature
LOCATION: 715
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AC002128
NAME/KEY: Misc_Feature
LOCATION: 1229
OTHER INFORMATION: diverging insertion, G in ref genbank:AC002128
NAME/KEY: Misc_Feature
LOCATION: 3676
OTHER INFORMATION: diverging nucleotide, T in ref genbank:AC002128
NAME/KEY: Misc_Feature
LOCATION: 5039
OTHER INFORMATION: diverging deletion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 5118
OTHER INFORMATION: diverging nucleotide, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 7337
OTHER INFORMATION: diverging deletion, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 8294
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 8604
OTHER INFORMATION: diverging nucleotide, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 8928
OTHER INFORMATION: diverging nucleotide, A in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 9021
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 9851
OTHER INFORMATION: diverging insertion, GATGAAA in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 9878
OTHER INFORMATION: diverging nucleotide, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 11475
OTHER INFORMATION: diverging nucleotide, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 11577
OTHER INFORMATION: diverging deletion, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 11775
OTHER INFORMATION: diverging nucleotide, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 13411
OTHER INFORMATION: diverging deletion, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 13538
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 13896
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 14912
OTHER INFORMATION: diverging nucleotide, A in ref genbank:AD000684

```

```

NAME/KEY: Misc_Feature
LOCATION: 16732
OTHER INFORMATION: diverging nucleotide, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 17169
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 18946
OTHER INFORMATION: diverging deletion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 19474
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 20500
OTHER INFORMATION: diverging deletion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 20501
OTHER INFORMATION: diverging deletion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 20502
OTHER INFORMATION: diverging deletion, A in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21270
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21356
OTHER INFORMATION: diverging insertion, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21476
OTHER INFORMATION: diverging nucleotide, A in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21588
OTHER INFORMATION: diverging insertion, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21601
OTHER INFORMATION: diverging deletion, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21635
OTHER INFORMATION: diverging insertion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 19963..19965
OTHER INFORMATION: Potential variant splicing site AAC
NAME/KEY: Misc_Feature
LOCATION: 1..21721
OTHER INFORMATION: n = a,g,c or t
US-09-269-939A-41

```

```

Query Match 3.6%; Score 39.2; DB 4; Length 21721;
Best Local Similarity 54.1%; Pred. No. 1.7;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

```

QY 31 GGCACAGGCGCGCCGAGATGCGCGCGCGCGCGCGCTTGCGGACGCGCTTCTGCTG 90
DB 21327 GGGCGGGGCGCAGTGGCGACGCGCTTGCGCGCGCGCGCGCGCGCGCTG 21268
QY 91 CTGGCGCGTGGCGCTGCTTGGCTTCGCGGCGGCGCGCGCGCGCGCGCGCGCGAT 150
DB 21267 CTGGCGCGGAGACGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 21208
QY 151 CCCCTCCCGGCGCACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 178
DB 21207 TTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21180

```

```

RESULT 12
US-09-269-939A-19/c
Sequence 19, Application US/09269939A
Patent No. 6635431
GENERAL INFORMATION:
APPLICANT: Bihain, Bernard
APPLICANT: Bougueleret, Lydie
APPLICANT: Yen-Potin, Frances
TITLE OF INVENTION: LSR Receptor, Its Activity, Its Cloning and Its Application To

```

```

1 EARLIER FILING DATE: 1999-07-20
2 NUMBER OF SEQ ID NOS: 20
3 SOFTWARE: Patent.pm
4 SEQ ID NO 1
5 LENGTH: 23187
6
7 TYPE: DNA
8 ORGANISM: Homo sapiens
9 FEATURE:
10 NAME/KEY: exon
11 LOCATION: 3540..3884
12 OTHER INFORMATION: exon2
13 FEATURE:
14 NAME/KEY: exon
15 LOCATION: 12163..12282
16 OTHER INFORMATION: exon3
17 FEATURE:
18 NAME/KEY: exon
19 LOCATION: 15144..15200
20 OTHER INFORMATION: exon4
21 FEATURE:
22 NAME/KEY: exon
23 LOCATION: 15765..15911
24 OTHER INFORMATION: exon5
25 FEATURE:
26 NAME/KEY: exon
27 LOCATION: 19579..19752
28 OTHER INFORMATION: exon6
29 FEATURE:
30 NAME/KEY: exon
31 LOCATION: 19899..19958
32 OTHER INFORMATION: exon7
33 FEATURE:
34 NAME/KEY: exon
35 LOCATION: 20329..20957
36 OTHER INFORMATION: exon9
37 FEATURE:
38 NAME/KEY: exon
39 LOCATION: 21047..21187
40 OTHER INFORMATION: exon10
41 FEATURE:
42 NAME/KEY: polyA_signal
43 LOCATION: 21166..21173
44 OTHER INFORMATION: AATAAA
45 FEATURE:
46 NAME/KEY: misc_feature
47 LOCATION: 1..2000
48 OTHER INFORMATION: potential 5'regulatory region
49 FEATURE:
50 NAME/KEY: misc_feature
51 LOCATION: 22324..23187
52 OTHER INFORMATION: homology with USF2 gene in ref: emb1 Y076641
53 FEATURE:
54 NAME/KEY: primer_bind
55 LOCATION: 523..544
56 OTHER INFORMATION: upstream amplification primer 17-2
57 FEATURE:
58 NAME/KEY: primer_bind
59 LOCATION: 1047..1168
60 OTHER INFORMATION: downstream amplification primer 17-2, complement
61 FEATURE:
62 NAME/KEY: primer_bind
63 LOCATION: 946..963
64 OTHER INFORMATION: upstream amplification primer 99-4576
65 FEATURE:
66 NAME/KEY: primer_bind

```

LOCATION: 1385..1402
OTHER INFORMATION: downstream amplification primer 99-4576 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 1096..1115
OTHER INFORMATION: upstream amplification primer 9-19
FEATURE:
NAME/KEY: primer bind
LOCATION: 1616..1635
OTHER INFORMATION: downstream amplification primer 9-19 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 1602..1621
OTHER INFORMATION: upstream amplification primer 9-20
FEATURE:
NAME/KEY: primer bind
LOCATION: 2074..2093
OTHER INFORMATION: downstream amplification primer 9-20 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 2036..2053
OTHER INFORMATION: upstream amplification primer 99-4557
FEATURE:
NAME/KEY: primer bind
LOCATION: 2563..2580
OTHER INFORMATION: downstream amplification primer 99-4557 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 2084..2102
OTHER INFORMATION: upstream amplification primer 9-1
FEATURE:
NAME/KEY: primer bind
LOCATION: 2483..2500
OTHER INFORMATION: downstream amplification primer 9-1 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 2470..2489
OTHER INFORMATION: upstream amplification primer 9-21 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 2062..2081
OTHER INFORMATION: downstream amplification primer 9-21
FEATURE:
NAME/KEY: primer bind
LOCATION: 3455..3474
OTHER INFORMATION: upstream amplification primer 9-3
FEATURE:
NAME/KEY: primer bind
LOCATION: 3882..3901
OTHER INFORMATION: downstream amplification primer 9-3 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 3775..3792
OTHER INFORMATION: upstream amplification primer 99-4558
FEATURE:
NAME/KEY: primer bind
LOCATION: 4336..4356
OTHER INFORMATION: downstream amplification primer 99-4558 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 4902..4920
OTHER INFORMATION: upstream amplification primer 99-14419 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 4444..4463
OTHER INFORMATION: downstream amplification primer 99-14419
FEATURE:
NAME/KEY: primer bind
LOCATION: 6638..6655
OTHER INFORMATION: upstream amplification primer 99-4577
FEATURE:
NAME/KEY: primer bind
LOCATION: 7072..7089

OTHER INFORMATION: downstream amplification primer 99-4577 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 7995..8012
OTHER INFORMATION: upstream amplification primer 99-4559
FEATURE:
NAME/KEY: primer bind
LOCATION: 8576..8593
OTHER INFORMATION: downstream amplification primer 99-4559 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 9622..9639
OTHER INFORMATION: upstream amplification primer 99-3148
FEATURE:
NAME/KEY: primer bind
LOCATION: 10023..10040
OTHER INFORMATION: downstream amplification primer 99-3148 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 9964..9981
OTHER INFORMATION: upstream amplification primer 99-4560
FEATURE:
NAME/KEY: primer bind
LOCATION: 10996..11015
OTHER INFORMATION: upstream amplification primer 99-14411 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 10546..10563
OTHER INFORMATION: downstream amplification primer 99-4560 , complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 10492..10512
OTHER INFORMATION: downstream amplification primer 99-14411
FEATURE:
NAME/KEY: primer bind
LOCATION: 11972..11990
OTHER INFORMATION: upstream amplification primer 99-4561
FEATURE:
NAME/KEY: primer bind
LOCATION: 12481..12501
OTHER INFORMATION: downstream amplification primer 99-4561 , complement
Query Match 3.6%; Score 39.2; DB 4; Length 23187;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 31 GGCACAGGGGCCGAGATGCGGCGGGCGGGCGGCTCGGGACGCGCTTCTGTG 90
DB 21420 GGGCCGGGGCGCATGCGACGCTGCGGGCGGGCGGGCGGACCGACGCGGGCGGTG 21361
QY 91 CTGGGGCTGGGCGCTTGGCTCGGGTGTTCGGGCTGCTCCGGGCTCGGGCAT 150
DB 21360 CTGGGGGGGACAGTCTCCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCTGCTC 21301
QY 151 CCCCTCCGGCGCGCACTGCGCGGTGCC 178
DB 21300 TTCCTCCGGCGCGCACTCCGGTCCC 21273
RESULT 14
US-09-252-991A-10838/c
; Sequence 10838, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10838
 ; LENGTH: 429
 ; TYPE: DNA
 ; ORGANISM: *Pseudomonas aeruginosa*
 ;
 US-09-252-991A-10838

Query Match	3.6%	Score 39;	DB 4;	Length 429;
Best Local Similarity	51.4%;	Pred. No. 0.19;		
Matches 90;	Conservative 0;	Mismatches 85;	Indels 0;	Gaps 0

<i>QY</i>	5	TCCGACACAGGGCGAGGATAGACACAGCAGGGCGGCCAGATATCGGCCTGGGGCGCGGC	64
<i>Db</i>	314	TCGCCGCGCTGGGGCGGGCGTCAACCGAGAGATGTGGACGTCCGACAAAGGCTCGGTCTCGACG	258
<i>QY</i>	65	GCGGCTCCGGGSAAGCGCCTTCCGTGATCGTAGGGCGCTGGCGCTTTTGCTCGCGATGGTT	124
<i>Db</i>	254	GCTGGGCTTTTGTTCGGGCGCCTGGGCGCTTGGCTTCGGGAGCTCAATCTCTTGGCGGCGTGG	195
<i>QY</i>	125	C CGGGCTGGTCCCGGCTCTCGGGAATCCCTCTCCGAGCCCACTGGACGGATGCCCT	179
<i>Db</i>	194	CGGTATTGCTGCTGGCGTTGGCTTGGGAGGCGCGGCTGCTGTGGCGCGGCGCT	140

```

RESULT 15
US-09-252-991A-10890
/ Sequence: 10890, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS.
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252.991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 10890
/ LENGTH: 474
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-10890

```

Query Match	3.6%;	Score 39;	DB 4;	Length 474;
Best Local Similarity	51.4%;	Pred. No. 0.21;		
Matches 90;	Conservative 0;	Mismatches 85;	Indels 0;	Gaps 0

2y 5 TCGGACACGAGGCGACAGAGTGTGACACGCGACACAGGCGCGCCGAGATATCGCGCGGAGCGCGGCG 64
Db 176 TCGCGCGCGTGGGCGGCGCTCAGCGCGCGAGATGTGGCGTGGCCGACAAAGCGGTTCTCGGCG 235
2y 65 GCGGCTCGGAGACGCGGCTTCTCTGAGTGTGGGCGCGCTGGCGCTGTGGCTGCGGATGTT 124
2b 236 GCTGGGCTGTGTGCGGCGCGCTCGGCGCTCGCGCTCGCATGTCTTCCGCGCGCTGG 295
2y 125 CCGGCGTGTCTCCGCGGTCTCGGCGCATCCCTCTCCGCGCGCACTGACCGAGTCCCT 179
2b 296 CGGTATGTGTGTGCGTGTGCTTGGCGCGCGCGCGCTGTCTTTCGCGCGCGCT 350

```
Search completed: April 22, 2004, 04:42:56
Job time : 110.289 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 00:14:11 ; Search time 482.252 Seconds
(without alignments)
9549.047 Million cells updates/sec

Title: US-10-010-050a-1_COPY_1_1084

Perfect score: 1084

Sequence: 1 gattccggcagcagggcagg.....acaaacactctctgttta 1084

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: Geneseqn299a04:*
2: geneseqn19808:*
3: geneseqn19908:*
4: geneseqn20008:*
5: geneseqn20018:*
6: geneseqn20028:*
7: geneseqn20038:*
8: geneseqn20048:*
9: geneseqn20058:*
10: geneseqn20068:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1084	100.0	1486	2	AAx02855 Human zsi
2	1067	98.4	1751	2	AAz24826 Human sec
3	1064.6	98.2	2120	4	AAK94829 Human ful
4	772.4	71.3	1038	2	AAx02866 Human deg
5	567	52.3	697	4	AAK93424 Human ctn
6	567	52.3	697	4	AAK92158 Human ctn
7	491	45.3	494	4	AAI25086 Probe #15
8	491	45.3	494	4	ABA70786 Human fce
9	491	45.3	494	4	AAI50961 Probe #19
10	491	45.3	494	4	ABA37272 Probe #15
11	491	45.3	494	4	AAK44992 Human bcn
12	491	45.3	494	4	AAK19040 Human bra
13	491	45.3	494	4	ABE44658 Human liv
14	491	45.3	494	6	ABE19237 Human gen
15	401.4	37.0	506	2	AAV88347 EST clone
16	380.2	35.1	474	4	AAI5879 Human fce
17	380.2	35.1	474	4	ABA58185 Human fce
18	380.2	35.1	474	4	AAI37794 Probe #64
19	380.2	35.1	474	4	ABA27366 Probe #58
20	380.2	35.1	474	4	AAK1927 Human bcn
21	380.2	35.1	474	4	AAK06266 Human bra
22	380.2	35.1	474	4	ABE31617 Human liv
23	380.2	35.1	474	6	ABE06689 Human gen

24	266	24.5	497	6	ABO55979 Human ova
25	197.2	18.2	351	6	AB179601 Human ova
26	133	12.3	473	6	AB166053 Lung cano
27	133	12.3	473	6	AB168165 Kidney ca
28	112.2	10.4	630	6	ABQ40653 Oligonuc1
29	112.2	10.4	630	6	ABQ40652 Oligonuc1
30	112.2	10.4	631	6	ABQ13702 Oligonuc1
31	112.2	10.4	631	6	ABQ13703 Oligonuc1
32	89.6	8.3	630	6	ABQ40651 Oligonuc1
33	89.6	8.3	631	6	ABQ40650 Oligonuc1
34	89.6	8.3	631	6	ABQ13704 Oligonuc1
35	89.6	8.3	631	6	ABQ13705 Oligonuc1
36	60	5.5	60	6	ABN37591 Human spl
37	46.8	4.3	2000	7	ADA71938 Rice gene
38	45.4	4.2	1945	6	AB217427 Arabidops
39	44.4	4.1	2000	7	ADA71938 Rice gene
40	44	4.1	1981	9	ADB87400 Transgene
41	43.8	4.0	17419	4	AA545393 Chemical1
42	43.8	4.0	17419	6	AB133295 Human imm
43	43.8	4.0	17419	6	ABK28238 DNA trans
44	43.4	4.0	4241	5	AB17545 Human ner
45	43.4	4.0	4241	5	AB17546 Human ner

ALIGNMENTS

RESULT 1	AAx02855	standard; DNA; 1466 BP.
ID	XX	AAx02855
AC	XX	AAx02855;
DT	XX	14-MAY-1999 (first entry)
DE	XX	Human zsig46 DNA.
KW	XX	Secreted protein; zsig46; human; chromosome 13; thyroid; disease;
KW	XX	hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;
KW	XX	Hirschprung's disease; neuronal ceroid-lipofucinosi; Wilson disease;
KW	XX	Reiger syndrome; immunoassay; detection; anti-idiotypic antibody;
KW	XX	therapy; diagnostic; ss.
OS	XX	Homo sapiens.
EH	XX	Key
FT	XX	CDS
FT	XX	Location/Qualifiers
FT	XX	47..1087
FT	XX	/*tag= a
FT	XX	/product= "zsig46"
XX	XX	W09905275-A1.
XX	XX	04-FEB-1999.
XX	XX	24-JUL-1998; 98WO-US015431.
XX	XX	24-JUL-1997; 97US-0053613P.
XX	XX	(ZYMO) ZYMOGENETICS INC.
XX	XX	Sheppard PO, Gilbertson DG;
XX	XX	WPI, 1999-14230/12.
XX	XX	P-PSDB; AAW92967.
XX	XX	New secreted polypeptide, zsig46, and its fragments, related fusion
XX	XX	proteins - used for diagnosis and treatment of thyroid disorders or
XX	XX	diseases involving genes on chromosome 13.
XX	XX	Claim 27; Page 88-90; 101pp; English.
XX	XX	This invention describes the isolation of a novel human secreted protein,
XX	XX	zsig46 encoded by a gene on chromosome 13 which is mainly expressed in

CC the thyroid. This product can be used to study secretion of proteins from
 CC cells and also to treat or prevent deficient expression of zsig46, which
 CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
 CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
 CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's
 CC syndrome). Antibodies and other binding proteins, are used as immunoassay
 CC reagents to detect zsig46 or cells expressing it, e.g. for assessing
 CC thyroid function to produce anti-idiotypic antibodies, for affinity
 CC purification of zsig46, to screen expression libraries, to neutralise
 CC zsig46 activity, and to deliver toxins, radioisotopes etc. for
 CC therapeutic or diagnostic purposes. Agonists of the product can be used
 CC to promote growth, differentiation and proliferation of specific cell
 CC types, e.g. for treating (extra)thyroid diseases or as additive to cell
 CC cultures
 CC
 SQ Sequence 1486 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;

Query Match 100.0%; Score 1084; DB 2; Length 1486;
 Best Local Similarity 100.0%; Pred. No. 2.8e-271;
 Matches 1084; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTGGGACGAGGGGCGAGAGTGAACACGCGACAGGCGCCGAGATGCGCGGCGCGC 60
 DB 1 GAATTGGGACGAGGGGCGAGAGTGAACACGCGACAGGCGCGCCGAGATGCGCGGCGCGC 60
 QY 61 GGGCGCGGCTGGGGGACGCGCTTCTGGTGGCTGGGGGCGCTGGCTTGGCTGGCGGT 120
 DB 61 GGGCGCGGCTGGGGGACGCGCTTCTGGTGGCTGGGGGCGCTGGCTTGGCTGGCGGT 120
 QY 121 GGTTCGGGGCTGGTCCGGGGTCTCGGGGATCCCTCCCGGCGCACTGGCGGCTCCCTA 180
 DB 121 GGTTCGGGGCTGGTCCGGGGTCTCGGGGATCCCTCCCGGCGCACTGGCGGCTCCCTA 180
 QY 221 GGTTCGGGGCTGGTCCGGGGTCTCGGGGATCCCTCCCGGCGCACTGGCGGCTCCCTA 180
 DB 221 GGTTCGGGGCTGGTCCGGGGTCTCGGGGATCCCTCCCGGCGCACTGGCGGCTCCCTA 180
 QY 181 CAGCGCTTTGACTTCGCTCCAAACCTGATCTTATTTGTAAGCTAATGTAATCTTCTG 240
 DB 181 CAGCGCTTTGACTTCGCTCCAAACCTGATCTTATTTGTAAGCTAATGTAATCTTCTG 240
 QY 241 TCCAACTGGCTCAGCTATCCAGTATGAGGGGATGATGATGATGATGATGATGATGATG 300
 DB 241 TCCAACTGGCTCAGCTATCCAGTATGAGGGGATGATGATGATGATGATGATGATGATG 300
 QY 301 ACAAGCCCCAGTATGGAATTTAATATATGAGACCTCTCGGAGCACTTGAATTTATGCA 360
 DB 301 ACAAGCCCCAGTATGGAATTTAATATATGAGACCTCTCGGAGCACTTGAATTTATGCA 360
 QY 361 TGAATGCCATTGGATTCAAGATGATTAAGTGGCAAGACTTACATGATGATGATGATG 420
 DB 361 TGAATGCCATTGGATTCAAGATGATTAAGTGGCAAGACTTACATGATGATGATGATG 420
 QY 421 ACTTTTCAACTGGCACTGATCTTCCCATCTCCGACTGCAATGATGATGATGATGATG 480
 DB 421 ACTTTTCAACTGGCACTGATCTTCCCATCTCCGACTGCAATGATGATGATGATGATG 480
 QY 481 CTGGTGTATCAAGCGCTGCTGCTTTTGAAGGAATGATGATGATGATGATGATGATG 540
 DB 481 CTGGTGTATCAAGCGCTGCTGCTTTTGAAGGAATGATGATGATGATGATGATGATG 540
 QY 541 AAATGGGACATTAGTTCAAGTGAAGCACTATATCAGGAAACATGTTCAACAAATGGCAA 600
 DB 541 AAATGGGACATTAGTTCAAGTGAAGCACTATATCAGGAAACATGTTCAACAAATGGCAA 600
 QY 601 GTGGGTGAAGCAGGACATGAAACAGAAATTTATATGAGATGATGATGATGATGATG 660
 DB 601 GTGGGTGAAGCAGGACATGAAACAGAAATTTATATGAGATGATGATGATGATGATG 660
 QY 661 CCGAGAAAAGGGGCGAGAGCATGGTTGATCTCAGAGCTGTTCCAAATTTGTTAG 720
 DB 661 CCGAGAAAAGGGGCGAGAGCATGGTTGATCTCAGAGCTGTTCCAAATTTGTTAG 720
 QY 721 GACCTTTAAAGATTTGGTGAATTTGGAGAGTTCAGAAACATGAAACCACTATAC 780
 DB 721 GACCTTTAAAGATTTGGTGAATTTGGAGAGTTCAGAAACATGAAACCACTATAC 780

QY 781 AAGATATTTCTTACAGTGGAGAACCTACTATCTGGGAAATGAACATCTGTTTTGG 840
 DB 781 AAGATATTTCTTACAGTGGAGAACCTACTATCTGGGAAATGAACATCTGTTTTGG 840
 QY 841 GCCACAGAAAACAAGACTCTGTTTACGATTAACAAAAGATTTATATACCTTCAAAAC 900
 DB 841 GCCACAGAAAACAAGACTCTGTTTACGATTAACAAAAGATTTATATACCTTCAAAAC 900
 QY 901 ACATTTGCCAATTAAGAAATTTCTGTTGAGTCTCTCGAAATTTTGTATGACGATTTGT 960
 DB 901 ACATTTGCCAATTAAGAAATTTCTGTTGAGTCTCTCGAAATTTTGTATGACGATTTGT 960
 QY 961 GCAAAACAGTTCATTTGTTTATATTTGAATATGTTTATGATGATGATGATGATGATG 1020
 DB 961 GCAAAACAGTTCATTTGTTTATATTTGAATATGTTTATGATGATGATGATGATGATG 1020
 QY 1021 TTTTATTAATAATACATATGAGAAATCCCTTACATGAGAAACAAACACTCTCTGG 1080
 DB 1021 TTTTATTAATAATACATATGAGAAATCCCTTACATGAGAAACAAACACTCTCTGG 1080
 QY 1081 TTTA 1084
 DB 1081 TTTA 1084

RESULT 2

ID AA224826 standard; DNA, 1751 BP.

AA224826;

02-DEC-1999 (first entry)

Human secreted protein gene 16 clone HMZADU77.

Human; secreted protein; fusion protein; gene therapy; protein therapy;
 diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 inflammatory; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

MO9947540-A1.

23-SEP-1999.

18-MAR-1999; 99MO-US005804.

19-MAR-1998; 98US-0078563P.
 19-MAR-1998; 98US-0078566P.
 19-MAR-1998; 98US-0078573P.
 19-MAR-1998; 98US-0078574P.
 19-MAR-1998; 98US-0078576P.
 19-MAR-1998; 98US-0078577P.
 19-MAR-1998; 98US-0078578P.
 19-MAR-1998; 98US-0078579P.
 19-MAR-1998; 98US-0078581P.
 01-APR-1998; 98US-0080312P.
 01-APR-1998; 98US-0080313P.
 01-APR-1998; 98US-0080314P.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
 Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;
 Shi Y, Moore PA.
 WPI; 1999-562050/47.

NR P-PSDB; AAY41323.

KX New isolated human genes, useful for diagnosis and treatment of e.g.
PT cancers, neurological disorders, immune diseases, inflammation or blood
PT disorders.

KX Claim 1; Page 308; 484pp; English.

CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. AA224802) for increasing the stability of the fused protein as
CC compared to the human protein only. The invention relates to 95 novel
CC genes and their fragments (nucleic acid sequences: AA224811-224907; amino
CC acid sequence AA1308-Y41404) which are useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,
CC pathological conditions can be diagnosed by determining the amount of the
CC new polypeptides in a sample or by determining the presence of mutations
CC in the new polynucleotides. Specific uses are described for each of the
CC 95 polynucleotides, based on which tissues they are most highly expressed
CC in (see AA224811 for described uses)

CC Sequence 1751 BP; 477 A; 355 C; 380 G; 523 T; 0 U; 16 Other;

CC Query Match 98.4%; Score 1067; DB 2; Length 1751;

CC Best Local Similarity 99.6%; Pred. No. 7.9e-267; Mismatches 2; Indels 0; Gaps 0;

CC Matches 1067; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

2Y 14 GGGGAGAGAGTACGACGCGCAAGGGCGCCGAGATGCGGCGGCGCGGCTTCG 73
2b 16 GCGGAGAGAGTACGACGCGCAAGGGCGCCGAGATGCGGCGGCGCGGCTTCG 75
2Y 74 GAGCGCGCTTCTGCTGCTGCGCGCTGCGGCTTCGCTTCGCTTCGCTTCG 133
2b 76 GAGCGCGCTTCTGCTGCTGCGCGCTGCGGCTTCGCTTCGCTTCGCTTCG 135
2Y 134 TCCCGGAGTCTCGGGGATCCCTCCGCGCGCACTGGCGCGGCTTCGCTTCG 193
2b 136 TCCCGGAGTCTCGGGGATCCCTCCGCGCGCACTGGCGCGGCTTCGCTTCG 195
2Y 194 TTCCGTCGCAAAACCTGATCCTTATGTCAGCTAAGTACTTCTGCTCACTG 253
2b 196 TTCCGTCGCAAAACCTGATCCTTATGTCAGCTAAGTACTTCTGCTCACTG 255
2Y 254 CCTATCCGAGTATGAGGAGGATGATGATGATGATGATGATGATGATGAT 313
2b 256 CCTATCCGAGTATGAGGAGGATGATGATGATGATGATGATGATGATGAT 315
2Y 314 TGGGAAATTAATATGAGAGACCTCTGAGAGACCTTGAATTTATGATGATG 373
2b 316 TGGGAAATTAATATGAGAGACCTCTGAGAGACCTTGAATTTATGATGATG 375
2Y 374 TTCAAGAGTACATTAATGAGAGACCTCTGAGAGACCTTGAATTTATGAT 433
2b 376 TTCAAGAGTACATTAATGAGAGACCTCTGAGAGACCTTGAATTTATGAT 435
2Y 434 GGCACACTGTAATTTCCCATCTCCGACCTGAAATGATGCCCTTTCTGCT 493
2b 436 GGCACACTGTAATTTCCCATCTCCGACCTGAAATGATGCCCTTTCTGCT 495
2Y 494 GGCAGTGTGCTGCTTTTGAAGGAGTATGATGATGATGATGATGATGAT 553
2b 496 GGCAGTGTGCTGCTTTTGAAGGAGTATGATGATGATGATGATGATGAT 555
2Y 554 GTTCAAGTACATTAATGAGAGACCTCTGAGAGACCTTGAATTTATGAT 613
2b 556 GTTCAAGTACATTAATGAGAGACCTCTGAGAGACCTTGAATTTATGAT 615
2Y 614 GACATATGAAACGAAATTTATATGAGAGATGATGATGATGATGATGAT 673
2b 616 GACATATGAAACGAAATTTATATGAGAGATGATGATGATGATGATGAT 675

```

```

QY 674 GCAGAGACATGTTGATTCCTACGACTGTCGCAATTTGTTAAGGACCTTAC 733
Db 676 GCAAGACATGTTGATTCCTACGACTGTCGCAATTTGTTAAGGACCTTAC 735
QY 734 TTGCTGAAATTTGAGAGAGAGTTCAGAAACATAGAAACCACTATACAG 793
Db 736 TTGCTGAAATTTGAGAGAGAGTTCAGAAACATAGAAACCACTATACAG 795
QY 794 TACGATGAGAACTTACTTATCTGAGAAATGAAACATCTGTTTTGGCC 853
Db 796 TACGATGAGAACTTACTTATCTGAGAAATGAAACATCTGTTTTGGCC 855
QY 854 AAGACTCTGTTAGCAATTAAGATTTATTAACCTTCAACCACTATTCGCA 913
Db 856 AAGACTCTGTTAGCAATTAAGATTTATTAACCTTCAACCACTATTCGCA 915
QY 914 AAGAAATTTCTGTTAGAGTCTCTTCGAAATTTTGAATGAGATTTG 973
Db 916 AAGAAATTTCTGTTAGAGTCTCTTCGAAATTTTGAATGAGATTTG 975
QY 974 TATTTGTTTATTAATTTGATATGTTTGTATGATGATGATGATGAT 1033
Db 976 TATTTGTTTATTAATTTGATATGTTTGTATGATGATGATGATGAT 1035
QY 1034 ACATATGAGAAATCCCTTACTATCAGAAACAAACACTCTGTTT 1084
Db 1036 ACATATGAGAAATCCCTTACTATCAGAAACAAACACTCTGTTT 1086

```

RESULT 3

ID AAK94829 standard; cDNA; 2120 BP.

AAK94829;

06-NOV-2001 (first entry)

Human full-length cDNA, SEQ ID NO: 3977.

Human, full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

EP1130094-A2.

05-SEP-2001.

07-JUL-2000; 2000EP-00114089.

08-JUL-1999; 99JP-00194486.

11-JAN-2000; 2000JP-00118774.

02-MAY-2000; 2000JP-00183765.

(HELI-) HELIX RES INST.

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

WPI; 2001-524255/58.

P-PSDB; AAM93870.

830 Primers useful for synthesizing full length cDNA clones and their use

in genetic manipulation.

Claim 8; SEQ ID NO 3977; 1380pp + Sequence Listing; English.

The invention relates to primers for synthesizing full length cDNA clones, 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesized by the oligo-capping method. The primers

CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO

XX Sequence 2120 BP, 590 A, 409 C, 460 G, 661 T, 0 U, 0 Other;

Query Match 98.2%; Score 1064.6; DB 4; Length 2120;
Best Local Similarity 99.6%; Pred. No. 3.6e-266;
Matches 1067; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 14 GGGGAGAGGATGACACGGGACAGAGGCGCCGAGATGCGGGCGGGGCGGGCGCTCGG 73
DB 19 GCGGAGAGGATGACACGGGACAGAGGCGCCGAGATGCGGGCGGGGCGGGCGCTCGG 78
QY 74 GGAGCGCGCTTCCTGCTGCTGCGGGCGCTGCGGGCGCTGCGGGCGCTGCGGGCGCTG 133
DB 79 GGAAGCGCTTCCTGCTGCTGCGGGCGCTGCGGGCGCTGCGGGCGCTGCGGGCGCTG 138
QY 134 TCCCGGGCTTCGGGCGATCCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTGAC 193
DB 139 TCCCGGGCTTCGGGCGATCCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTGAC 198
QY 194 TTCCGTCGAAACCTGATCCTTATTTGCAAGCTAAGTACTTTCTTCTGCTCACTGCTCA 253
DB 199 TTCCGTCGAAACCTGATCCTTATTTGCAAGCTAAGTACTTTCTTCTGCTCACTGCTCA 258
QY 254 CCATCCGAGTATGAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 313
DB 259 CCAATCCGAGTATGAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 318
QY 314 TGGGAAATTTAAATATGAGGACCTTCCTGCGGACACTTGGAAATTTATGATGATGATGAT 373
DB 319 TGGGAAATTTAAATATGAGGACCTTCCTGCGGACACTTGGAAATTTATGATGATGATGAT 378
QY 374 TTGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 433
DB 379 TTGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 438
QY 434 GGCAGCTGTACATTTCCCGATCTCCGACCTGAAATGATGATGATGATGATGATGATGAT 493
DB 439 GGCAGCTGTACATTTCCCGATCTCCGACCTGAAATGATGATGATGATGATGATGATGAT 498
QY 494 GGGCGCTGCTGCTTTTGAAGGAATGATGATGATGATGATGATGATGATGATGATGAT 553
DB 499 GGGCGCTGCTGCTTTTGAAGGAATGATGATGATGATGATGATGATGATGATGATGAT 558
QY 554 GTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 613
DB 559 GTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 618
QY 614 GACATGAAACAGGAATTTATGAGACATGAGATGATAAAGCCAGCCAGAAAAGGGG 673
DB 619 GACATGAAACAGGAATTTATGAGACATGAGATGATAAAGCCAGCCAGAAAAGGGG 678
QY 674 GCGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 733
DB 679 GCGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
QY 734 TTGGCTGGAATTTGAGAGAGGATTCAGAAACATGAGAAACCAATGACAAAGATTTTCTT 793
DB 739 TTGGCTGGAATTTGAGAGAGGATTCAGAAACATGAGAAACCAATGACAAAGATTTTCTT 798
QY 794 TACAGTGAAGACCTTATCTGAGAAATGAAACATCTGTTTTTGGGCGAAGAGAAAC 853
DB 799 TACAGTGAAGACCTTATCTGAGAAATGAAACATCTGTTTTTGGGCGAAGAGAAAC 858
QY 854 AAGACTCTGCTTTAGGCAATAAAAAGATTTTATTAATCCCTTCAACCACTTGGCAACT 913
DB 859 AAGACTCTGCTTTAGGCAATAAAAAGATTTTATTAATCCCTTCAACCACTTGGCAACT 918
QY 914 AAGACTCTGCTTTAGGCAATAAAAAGATTTTATTAATCCCTTCAACCACTTGGCAACT 973

```

```

DB 919 AAGAAATTTCTGTTGAGTCTCTGCAAAATTTTATGATGATGATGATGATGATGATGAT 978
QY 974 TATTTGTTTATATATTTTGAATATGATGATGATGATGATGATGATGATGATGATGAT 1033
DB 979 TATTTGTTTATATATTTTGAATATGATGATGATGATGATGATGATGATGATGATGAT 1038
QY 1034 ACATATGAAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGTTTA 1084
DB 1039 ACATATGAAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGTTTA 1089

```

```

RESULT 4
AAK02866
ID AAK02866 standard; DNA, 1038 BP.
XX
AC AAK02866;
XX
DT 14-MAY-1999 (first entry)
XX
DE Human degenerate zsig46 DNA.
XX
KM Secreted protein; zsig46; human; chromosome 13; thyroid disease;
KM hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;
KM Hirschsprung's disease; neuronal ceroid-lipofucinosi; Wilson disease;
KM Reiger syndrome; immunoassay; detection; anti-idiotypic antibody;
KM therapy; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN MO9905275-A1.
XX
PD 04-FEB-1999.
XX
PF 24-JUL-1998; 98MO-US015431.
XX
PR 24-JUL-1997; 97US-0053613P.
XX
PA (ZIMO) ZYMOGENETICS INC.
XX
PI Sheppard PO, Gilbertson DG;
XX
DR WPI; 1999-142930/12.
XX
PT New secreted polypeptide, zsig46, and its fragments, related fusion
PT proteins - used for diagnosis and treatment of thyroid disorders or
PT diseases involving genes on chromosome 13.
XX
PS Claim 31; Page 94-95; 101bp; English.
XX
CC This invention describes the isolation of a novel human secreted protein,
CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in
CC the thyroid. This product can be used to study secretion of proteins from
CC cells and also to treat or prevent deficient expression of zsig46, which
CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's
CC disease, neuronal ceroid-lipofucinosi, Wilson disease and Reiger
CC syndrome). Antibodies and other binding proteins, are used as immunoassay
CC reagents to detect zsig46 or cells expressing it, e.g. for assessing
CC thyroid function to produce anti-idiotypic antibodies, for affinity
CC purification of zsig46, to screen expression libraries, for neutralise
CC zsig46 activity, and to deliver toxins, radioisotopes etc. for
CC therapeutic or diagnostic purposes. Agonists of the product can be used
CC to promote growth, differentiation and proliferation of specific cell
CC types, e.g. for treating (extra)thyroid diseases or as additive to cell
CC cultures
XX
SQ Sequence 1038 BP, 189 A, 106 C, 179 G, 166 T, 0 U, 398 Other;

```

Query Match 71.3%; Score 772.4; DB 2; Length 1038;
Best Local Similarity 61.7%; Pred. No. 2.4e-190;
Matches 640; Conservative 232; Mismatches 165; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - nucleic search, using frame_plus_p2n model
on on: April 24, 2004, 22:59:32 ; Search time 6540.36 Seconds
(without alignments)
2292.945 Million cell updates/sec

Title: US-10-010-050A-2
Effect score: 1927
Sequence: 1 MRGAGAGRGRASWCNALAL.....IKITVEIPPIPRNKTLISGL 346

Coring table:
BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 3470272 seqs, 2167151695 residues
otal number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Out-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
MODEL=frame+ p2n.model -DEV=xlh
Q=/cgm2.1/USPTO.spool/US10010050/runat_22042004_113204_27549/app.query.fasta_1.1372
DB=GenBml -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCUT=0 -LOOPEXT=0
UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdt -LIST=45
DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=pro -NORW=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
USER=US10010050_@CGN_1.1_7433_@runat_22042004_113204_27549 -NCPU=6 -ICPU=3
NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELACK=100 -LONGLOG
DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: GenBml:*

1:	gb_ba:*
2:	gb_hcg:*
3:	gb_in:*
4:	gb_om:*
5:	gb_ov:*
6:	gb_pac:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_scs:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vl:*
15:	em_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_or:*
22:	em_ov:*
23:	em_pat:*
24:	em_ph:*
25:	em_pl:*
26:	em_ro:*
27:	em_scs:*
28:	em_un:*

29: em_vi:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pin:*

35: em_hcg_rod:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_sy:*

39: em_hugo_hum:*

40: em_hugo_mus:*

41: em_hugo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1927	100.0	1486	6 BD073402	BD073402 Secreted
2	1927	100.0	4080	9 AF068227	AF068227 Homo sapi
3	1916	99.4	1751	6 BD136338	BD136338 95 human
4	1913	99.3	2120	6 BD127858	BD127858 Primer fo
5	1913	99.3	2120	9 AK075109	AK075109 Homo sapi
6	1631	84.6	1038	6 BD073413	BD073413 Secreted
7	1451	75.3	2318	10 BC025487	BC025487 Mus muscu
8	1425	73.9	198524	2 AC109554	AC109554 Rattus no
9	1368	71.0	227920	2 AC107510	AC107510 Rattus no
10	1065	55.3	697	6 BD125187	BD125187 Primer fo
11	1065	55.3	697	6 BD126453	BD126453 primer fo
12	925.5	48.0	169362	9 AC001226	AC001226 Genomic s
13	925.5	48.0	251187	2 AL136440	AL136440 Homo sapi
14	925.5	48.0	251187	2 AL135987	AL135987 Homo sapi
15	730	37.9	110000	10 AB014175	AB014175 Mus muscu
16	730	37.9	156823	2 AC102815	AC102815 Mus muscu
17	727	37.7	506	6 BD060465	BD060465 Secreted
18	716	37.2	229015	2 AC106966	AC106966 Rattus no
19	711	36.9	217336	2 AC131344	AC131344 Rattus no
20	617.5	32.0	251187	2 AL135987	AL135987 Homo sapi
21	582	30.2	198524	2 AC109554	AC109554 Rattus no
22	272.5	14.1	258273	2 AC113855	AC113855 Rattus no
23	256	13.3	110000	3 AC116305	AC116305 Dictyoste
24	241	12.5	473	6 AX333881	AX333881 Sequence
25	241	12.5	473	6 AX335993	AX335993 Sequence
26	146.5	7.6	256879	3 AC116582	AC116582 Dictyoste
27	117	6.1	112027	2 AC007006	AC007006 Homo sapi
28	115	6.0	187727	2 AC064806	AC064806 Homo sapi
29	112	5.8	2162	2 BC014298	BC014298 Homo sapi
30	112	5.8	2175	6 AX881019	AX881019 Sequence
31	112	5.8	2175	6 BD158729	BD158729 Primer fo
32	112	5.8	2175	6 AK023964	AK023964 Homo sapi
33	111.5	5.8	124548	8 AP004083	AP004083 Oryza sat
34	111.5	5.8	223182	5 AL292949	AL292949 Zebrafish
35	110	5.7	236195	2 AC073713	AC073713 Mus muscu
36	109	5.7	37231	3 LMFL4803	LMFL4803
37	109	5.7	145500	2 AC137990	AC137990 Leishmani
38	108.5	5.6	263129	2 AC098289	AC098289 Rattus no
39	108	5.6	204545	2 AC108566	AC108566 Rattus no
40	106	5.5	110000	2 AL160016	AL160016 Continuation (4 of
41	106	5.5	130244	9 AL590705	AL590705 Human DNA
42	105	5.4	155156	9 AF241727	AF241727 Homo sapi
43	104.5	5.4	143847	2 AF205589	AF205589 Homo sapi
44	104.5	5.4	191078	2 AC022505	AC022505 Homo sapi
45	104	5.4	2061	9 HSM805115	AL833837 Homo sapi

ALIGNMENTS

RESULT 1

LOCUS	BD073402	1486 bp	DNA	linear	PAT 27-AUG-2002
DEFINITION	Secreted protein which human chromosome 13 encodes.				
ACCESSION	BD073402				
VERSION	BD073402.1	GI:22619005			
KEYWORDS	JP 2001511345-A/1.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	Sheppard,P.O. and Gilberton,D.G.				
TITLE	Secreted protein which human chromosome 13 encodes				
JOURNAL	Patent: JP 2001511345-A 1 14-AUG-2001; ZYMOGENETICS INC				
COMMENT	OS Homo sapiens (human) PN JP 2001511345-A/1 PD 14-AUG-2001 PF 24-JUL-1998 JP 2000504249 PR 24-JUL-1997 US 60/053613 PI PAUL O SHEPPARD, DIBRA G GILBERTON PC C12N1/09,A61K38/00,A61K48/00,C07K14/47,C07K16/18,C12N1/15, PC C12N1/19, PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,C12N15/00,A61K37/02,C12N5/10				
FEATURES	CC Secreted protein which human chromosome 13 encodes FH Key				
source	FT CDS	Location/Qualifiers	(47)..(1084).		
		1..1486			
		/organism="Homo sapiens"			
		/mol_type="genomic DNA"			
		/db_xref="taxon:9606"			
ORIGIN					
Alignment Scores:					
Pred. No.:	1,41e-166	Length:	1486		
Score:	1927.00	Matches:	346		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	6	Gaps:	0		
US-10-010-050A-2 (1-346) x BD073402 (1-1486)					
OY	1	MeTaAGaAGGjYAlaGjYAlaAlaAlaArgjYArGjYAsErTPCySTPaLeuAlaLeu	20		
Db	47	ATGCGCGCGGGGGCGCGCGCGCTCGGGGAAGCGGCTTCGTGCTGGGCGCTGGCGCTG	106		
OY	21	LeuTPLeuAlaAlaAlaAlaAlaProGjYTPSeArGjYAlSeRgYjYLeProSeArGArGhIs	40		
Db	107	CTTGGGCTCGGGGAGTGTTCGGGGCGGTCCGGGGTCTCGGGCATCCCTCCGGGCGCAC	166		
OY	41	TrpProValProTyrLysArGpPheAspPheArGpPolysProAspProTyrCyGlnAla	60		
Db	167	TGGCGCGGCGCCCTTCAGACGCGCTTGACATCCGATCCCAAAACCTGATCCCTATTGTCAAGCT	226		
OY	61	LysTyrTrpPheCySProThnGjYSerProLeProValMetGjYjYAspAspAspIle	80		
Db	227	AAGTAACTCTTCTGTCCAACTGGGCTCACCTATCCAGTTATGAGAGGATGATGACATT	286		
OY	81	GluValPheArGLeuGlnAlaProValTrpGjYpHeLysTyrGjYAspLeuLeuGjYHIs	100		
Db	287	GAAGTTTTCGATTACAAAGCCCAAGTATGGGAATTTAAATATGAGAGACTCTCGGACAC	346		
Db	347	TTGAATAATTAGCATGATGCATTCGATTCAGAAAGTACATTAACATGCGAAGAACTACACA	406		
OY	101	LeuLysIleMetHIsAspAlaIleGjYpHeArGSeThrLeuThrGjYLysAsnTyrThr	120		
Db	121	MetGjYTPYrGjYLeuPheGjYLeuGjYAsnCySThrPheProHIsLeuArGProGjY	140		
Db	407	ATGGAATGATAGAACTTTCACACTTGGCAACGTGATACATTTCCCATCTTCGACTGAA	466		

QY	141	Metrsplaprophe	TPCYSAANGINGYALALaCyshphePheguigYlileasp	160
Db	467	ATGATGACCCCTTTCTGGTGAATCAAGGCGCTCCGCTTTTGTGGAGAAATGATGAT		526
QY	161	ValHisTpyugl	uaenglyThrlieuAlGlnValAlaThrIleSerGlyAsnMetPhe	180
Db	527	GTTCACTGGAAGAAATGGAGCACTTAGTTCAAGTACACATATATACGAGAAACATGTTCC		586
QY	181	AsngImetAla	lySTpVallysgInaSpaengluThrglyIleYTYTgIuThTrp	200
Db	587	AACCAAAATGGCAAAAGTGGTGAAACAGAGACAAATGAAACAGAAATTTATGACACATGG		646
QY	201	AsnVallysaIase	rProGluYsgYAlaGluThrTrpPheAspSerTYrAspCysSer	220
Db	647	AATGTAAAGCCAGCCCGAAGAAAGGGGCGAGACATGGTTGATTCCTACGACTGTCC		706
QY	221	LysPheVal	leuArqThrPheAsnYsleuAlaGluPheGlyAlaGluPheYsAsnIle	240
Db	707	AAATTTGTGTTAAAGACCTTTAACAGTGGCTGAAATTTGGAGCAGAGTCCAAAGACATA		766
QY	241	GluThrsnYrThr	ArGlllePheLeuTYrSerGlyGluProThrTYrleuGlyAsnGlu	260
Db	767	GAAACCAACATATACMAAAATATTTCTTTACAGTGAAGAACCTATCTATCGGGAATGAA		826
QY	261	ThrservalPhe	gIyProThrGlyAsnYsleThrlieuGlyleuAlaIleYsArqPheTYr	280
Db	827	ACATCTGTTTTGGGCGACAGAGAAACAAACATCTTGTTAGCCATAAAAAGATTTAT		886
QY	281	TYrProPheYs	ProHisleuProThrLYsgIuPheleuIeuSerleuGlnIlePhe	300
Db	887	TACCCCTTCAAAACCACTTGGCCAACTAAAGAAATTTCTGTGAGTCTCTTGCAAAATTTT		946
QY	301	AspAlaValIle	ValHislysgInPheTYrleuPheYrAsnPheGluTYrTrpPheLeu	320
Db	947	GATGCACTGATGATGACAAACAGTCTATTTGTTATATATTTGAATATGTGTTTTTA		1006
QY	321	ProMetYsPhe	rPhePheIleYsIleThTYrGluGluIleProleuProIleArqAsn	340
Db	1007	CCATGAAATTCCTCTTATTAATAATACATATAGAAAAATCCCTTACTATACAGAAAC		1066
QY	341	LysThrlieu	SergIyeu 346	
Db	1067	AAAACACTCTCTGTTTA 1084		
RESULT 2				
AF068227				
LOCUS		4080 bp	mRNA	linear PRI 28-JUL-1998
DEFINITION		Homo sapiens putative transmembrane protein (CLN5)	mRNA, complete cds.	
ACCESSION	AF068227			
VERSION	AF068227.1	GI:3342385		
KEYWORDS				
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS		1 (bases 1 to 4080)		
TITLE		Santukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L.		
		CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis		
JOURNAL		Nat. Genet. 19 (3), 286-288 (1998)		
MEDLINE		98324783		
PubMed		9662406		
REFERENCE		2 (bases 1 to 4080)		
AUTHORS		Santukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L.		
TITLE		Direct Submission		
JOURNAL		Submitted (26-MAY-1998) Department of Human Molecular Genetics, National Public Health Institute, Mannerheimintie 166, Helsinki 00300, Finland		


```

566,
PC C12N15/00,A61K37/02,C12N5/00
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
FH Key location/Qualifiers
FH 1. .1751
FT source location/Qualifiers
FT 1. .1751
FEATURES
source location/Qualifiers
1. .1751
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

Alignment Scores:	
Pred. No.:	1,77e-165
Score:	195.00
Percent Similarity:	99.42%
Best Local Similarity:	99.42%
Query Match:	99.43%
DB:	6
Gaps:	0
Length:	175
Matches:	344
Conservative:	0
Mismatches:	2
Indels:	0
Gaps:	0

US-10-010-050A-2 (1-346) X BD136338 (1-1751)

QY	2	MeAaArgGIyAlaaglYAlaAlaArgGlyArArAlaSerTPCySTPAlaAlaAlaLeu	20
Db	49	ATCGCGCGGGGCGCGGGCGCGCTGGGGAGCGCGCTTCTGGCTGGCGCTGGCGCTG	108
ZY	21	LeuTPLeuAlaValAlaProGlyTPSeArGValSerGylleProSeArGArGHis	40
Db	109	CTTTGGCTCGGGTGGTTCCGGGCTGATCCCGGGCTCGGGCAATCCCTCCGGCGCAC	168
ZY	41	TrpProValProTyrTlyArGrpPheArGrpPolysProArApproTyrCySGlnAla	60
Db	169	TGGCGGGTCCACAAAGGCTTTTACCTTCGTCACAAACCTGATCTTATGTCAAGCT	228
QY	61	LysTyrThrPheCySProThrGlySerProLleProValIleMcGIuGlyAspAspAspIle	80
Db	229	AAGTAATCTTCTGTCCAACTGGCTCACTATCCAGATTATGAGGGGTATGATGACATT	288
QY	81	GluValPheArGrLeuGlnAlaProValTPrgIuPheLysTyrGlyAspLeuLeuGlyHis	100
Db	289	GAAGTTTTTCGATTACAAAGCCAGATAGGGGATTTAAATATGAGAGCTTCCTGGGACAC	348
QY	101	LeuLysIleMetHisAspAlaIleGlyPheArGSerTrleuThrGlyLysAsnTyrThr	120
Db	349	TTGAAATTTATGCAATATGCATTGGATTCAAGATCAATTAACTGGCAGAACTTACAA	408
QY	121	MetGluTPTrpGluPhePheGlnLeuGlyAsnCySThrPheProHisLeuArGrProGlu	140
Db	409	ATGGAATGATATGACTTTTCCACTTGGCACTGACATTTCCCATCTCCGACCTGAA	468
QY	141	MetAspAlaProPheTrpCyAsnGlnGlyAlaAlaCyPhePheGlnGlyIleAspAsp	160
Db	469	ATGATATCCCTTCTGTGATATCAAGGCGCTGCTCTTTTGAAGGAAATTAATGAT	528
QY	161	ValHisTrpLysGluPheGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPhe	180
Db	529	GTTCACTGGAAAGGAAATGGGACATTAGTTCAAGTAGCAACTATATCGAAACATGCTC	588
QY	181	AsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyrTyrGluThrTrp	200
Db	589	AACCAATATGCAAAAGTGGTGAAACCGACATGAACACGGAATTTATTAAGAAGCATGG	648
QY	201	AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSer	220
Db	649	AATATTAATAGCGCCGACGAAAGGGGGCGAGACATGGTTGATTCTCTCAACATGTTCC	708
QY	221	LysPheValLeuArGrThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle	240

Db	709	AAATTGTGTTAAGACCTTTAAACAAGTTGGCTGAATTTGGACACAGAGTTCAAGAACATA	768
Qy	241	GLuThrasnYrthrAgliLePheLeuYrSerGlyGluProthrYrLeuGlyAsnGlu	260
Db	769	GAAMCCACACTATACAGAAATATTTCTTTACAGGAGAACCTTACTATCTGGGAAAGAA	828
Qy	261	ThisSerValPheGlyProThrGlyAsnLeuYrThrLeuGlyLeuAlaIleYrArgPheTyr	280
Db	829	ACATCTGTTTTTGGGCCAACAGGAACAAACACTCTGGTTTAGCCATMAAAAAAGATTTTAT	888
Qy	281	TyrProPheLeuYrProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePhe	300
Db	889	TACCCCTTCAACACACACTTGGCCAACTMAAGAATTTCTGTGAGTCTCTGGCAAAATTTT	948
Qy	301	AspAlaValIleValHisLeuGlnPheTyrLeuPheTyrAsnPheGlyLysTyrTyrPheLeu	320
Db	949	GATGCAAGTGAATTGTGCACAAACAGTTCTATTTGTTTATATATTTTGAATATTTGTTT	1008
Qy	321	ProMetLysPheProPheIleLeuYsIleThrTyrGluGlnIleProLeuProIleArgAsn	340
Db	1009	CCATGAAATTCCTTTTATTTAAATAATACATATGAAAGAAATCCCTTTACCATCAGAAAC	1068
Qy	341	LysThrLeuSerGlyLeu 346	
Db	1069	AAAAACCTCTGGTTTA 1086	

RESULT 4	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE
BD127858	BD127858	Primer for synthesizing full-length cDNA and use thereof.	GI:23222803	JP 2002017375-A/3289.	Homo sapiens (human)

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2120)
 Oca, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 REFERENCE
 AUTHORS

TITLE	Primer for synthesizing full-length cDNA and use thereof
JOURNAL	Patent: JP 2002017375-A 3289 22-JAN-2002;
COMMENT	HELIX RESEARCH INSTITUTE OS Homo sapiens (human) acc no. U00608.1(1990)

PF 07-JUL--2000 JP 200025312
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA

```

1212N15/09,C07K16/47,C07K16/18,1212N1/15,1212N1/19,1212N1/21,1212N5/10,
10,
PC, 1212P1/02,1212P1/68/1212P1/08,G06F17/30,1212N15/00,1212N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS Location/Qualifiers
1..2130
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

Alignment Scores:	
Pred. No.:	4,26e-165
Score:	193.00
Percent Similarity:	99.42%
Best Local Similarity:	99.42%
	Length: 2120
	Matches: 344
	Conservative: 0
	Mismatches: 2

```

nery Match:      99.27%      Indels:      0
ID:              6          Gaps:          0
IS-10-010-050a-2 (1-346) x BD127858 (1-2120)

Y      1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
b      52 ATCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 111
Y      21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyLeuProSerArgGHis 40
b      112 CTTGGCTCGCGGCTGGTCCGGGCTGGTCCGGGCTGGGCGATCCCTCCCGGCCAC 171
Y      41 TrpProValProTyrLysArgPheAspPheArgProLysProAspProTyrCysGlnAla 60
b      172 TGGCCGGTGGCTCGCAAGCGCTTGACCTTCCTCCCAAAACCTGATCCTATTGTCAAGCT 231
Y      61 LysTrpTrpPheCysProTrpGlySerProLysProLysProValMetGlyLysAspAspAla 80
b      232 AAGTATACCTTTCTGTCCAACTGGCTCACTATCCAGTTATGGAGGGTGTATGACACT 291
Y      81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTrpGlyAspLeuGlyHis 100
b      292 GAAGTTTTCGATTACAGCCCAAGATGGGAAATTTAAATATGGAGACCTCCCGGACAC 351
Y      101 LeuLysTrpLeuMetHisAspAlaAlaGlyPheArgSerTrpLeuThrGlyLysAsnTyrThr 120
b      352 TTGAAAATTATGCAATATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 411
Y      121 MetGluTrpTrpGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGlu 140
b      412 ATGGAATGATATACCTTTCCAACTTGGCACTGACATTTCCCATCTCCCACTCGAA 471
Y      141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGlyLysAlaAsp 160
b      472 ATGGAATGCGCTTTCGAGTATCAAGGCGCTGCTTTTGAAGGAATGATGAT 531
Y      161 ValHisTrpLysGlnAsnGlyThrLeuValGlnValAlaThrLysSerGlyAsnMetPhe 180
b      532 GTTCACCTGGAAAGAAATGGAGCAATGATGCAATGCAATGCAATGCAATGCAATGCAAT 591
Y      181 AsnGluMetAlaLysTrpValLysGlnAspAsnGluThrGlyLysTrpGluThrTrp 200
b      592 AACCAATGGCAAGGCGGGAACAGCAACAAAGAAACGGAATTTATATGACATGAG 651
Y      201 AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTrpAspCysSer 220
b      652 AATGTAAGGCGCGCCAGAAAGGGGCGAGACATGCTTGATTCCTACGACTGCTTCC 711
Y      221 LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnLe 240
b      712 AATTTGTGTAAAGACCTTTACAAAGTTGGCTGAATTTGGAGCAAGATTCAAGAACATA 771
Y      241 GluThrAsnTrpThrArgLysPheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGlu 260
b      772 GAACCACTATACAAAGAAATTTCTTACAGGGAACCTACTATATGTGGAAATGAA 831
Y      261 ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaLysArgPheTyr 280
b      832 AACTGCTTTTGGCGCAACAGAAACAGACTCTTGCTTGGCCATAAAAGATTAT 891
Y      281 TyrProPheLysProHisLeuProThrLysGluPheLeuLysSerLeuLeuGlnLePhe 300
b      892 TACCCCTTAAACCAACATTTGGCACTAAAGAAATTTCTGTGAGTCTCTGCAAAATTTT 951
Y      301 AspAlaValLysValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu 320
b      952 GATGCGATGATTTGTGCAACAAGCTTCTATTGTTTATATTTTGAATATGGATTTT 1011
Y      321 PrometLysPheProPheLysLysLeuTyrGluGluLeuProLeuProLysArgAsn 340
b      1012 CCTATGAATTCCTTTTATTAATATACATATGAAGAAATCCCTTACTTACATCAAGAAC 1071

```

```

QY      341 LysThrLeuSerGlyLeu 346
b      1072 AAAACACTCTCTGATT 1089

RESULT 5
AK075109
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 Isogai,T., Ota,T., Nishikawa,T., Hayaashi,K., Otsubuki,T.,
Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
Kojima,S., Nagahara,K., Masuko,Y., Ono,T., Okano,K., Yoshikawa,Y.,
Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
Ninomiya,K.
NEDO human cDNA sequencing project
JOURNAL
Unpublished
2 (bases 1 to 2120)
REFERENCE
Isogai,T. and Otsubuki,T.
Direct Submission
JOURNAL
Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
Institute of Medical Science, University of Tokyo, Laboratory of
Genome Structure, Human Genome Center; cDNA 5'-& 3'-end one pass
sequencing and clone selection: Helix Research Institute (supported
by Japan Key Technology Center etc.).
FEATURES
source
1..2120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1003407"
/tissue_type="Placenta"
/clone_lib="PLACE1"
/note="Cloning vector: pME18SFL3"

ORIGIN
Alignment Scores:
Pred. No.: 4,266-165 Length: 2120
Score: 1913.00 Matches: 344
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 2
Query Match: 99.27% Indels: 0
Gaps: 0
DB: 9

US-10-010-050a-2 (1-346) x AK075109 (1-2120)
QY      1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
b      52 ATCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 111
Y      21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyLeuProSerArgGHis 40
b      112 CTTGGCTCGCGGCTGGTCCGGGCTGGTCCGGGCTGGGCGATCCCTCCCGGCCAC 171
Y      41 TrpProValProTyrLysArgPheAspPheArgProLysProAspProTyrCysGlnAla 60
b      172 TGGCCGGTGGCTCGCAAGCGCTTGACCTTCCTCCCAAAACCTGATCCTATTGTCAAGCT 231
Y      61 LysTrpTrpPheCysProTrpGlySerProLysProLysProValMetGlyLysAspAspAla 80

```

b		232	AAGTAATACCTTCTGTCCAACTGGCTCAACTATCCAGTTATTGAGGGTGATGATGAACA	291
2y		81	GluValPheArgLeuGlnAlaProValTrpGluPheLysTyrglyAspLeuLeuGlyHis	100
b		292	GAAAGTTTTTGATTTACAGGCCCAAGTATGGGAATTTAAAATATGAGACCTTCGGGAC	351
2y		101	LeuIleYlleMetHisAspAlaIleIleGlyPheArgSerThrLeuArgglyAsnTyrrThr	120
b		352	TTGAAATTTATGCATGATGCGCATTTGATTTCAAGAATTCATTAACTGGCAAGATCAC	411
2y		121	MetGluTrpTyrgluLeuPheGlnLeuGlyAsnCysThrPheProHisleuArgProGlu	140
b		412	ATGGAATAGATATGAACTTTTCCAACCTTGGCAACTGTGACATTTCCCACATCCGAC	471
2y		141	MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGlyIleAspAsp	160
b		472	ATGATATCCCCCTTCTGATGTAACAAGGGCGTCCGCTTTTGTGAGGAATTAGAT	531
2y		161	ValHisTrpIlysgIuAsngIyThrLeuValGlnValAlathrlIeserGlyAsnMetPhe	180
b		532	GTTCACTGGAAGGAATAATGGGACATTAGTTCAAATGACCACTAATATCAGGAACATG	591
2y		181	AsnGluMetAlalystrpValIlysgIuAspAsnGluThrGlyIleTyrrtyrGluThrTrp	200
b		592	AACCAATATGGCAAGTGGGTGAAAAACAGAACATGAAAAACGAATTTATATGAAACATG	651
2y		201	AsnValIylsalaserProGluIlysgIyAlaGluThrTrpPheAspSerTyraPCysSer	220
b		652	AATGTAAATAGCCAGCCACGAAAAAGGGGCGAGACATGGTTGATTCCTAGCATGTGCC	711
2y		221	LysPheValIeuArgThrPheAsnIlyseuAlacluPheGlyAlaGluPheIlyAsnIle	240
b		712	AAATTTGTGTTAAAGACCTTTTAAACAGTTGGCTCAATTTGGACGAGGTTCAGAACATA	771
2y		241	GluThrAsnTyrrThraArgIlePheLeuTyserGlyGluProTrpTyrrLeuGlyAsnGlu	260
b		772	GAAACCAACTATACAGAAATATTTCTTTACAGTGGAGAACCTTATCTGGGAAATGAA	831
Qy		261	ThisSerValPheglYProthrGlyAsnIlysthrLeugIyleuAlaileYsarPheTy	260
b		832	ACATCTGTTTTTGGGCCAACAGGAACAAAGACTCTTGGTTAGCCATVAAAAGATTTTAT	891
Qy		281	TyrTrpPheLysProHisleuProThrylIeGluPheLeuSerLeuLeuGlnIlePhe	300
b		892	TACCCTTCCAACACACATTTGCCAACPAAATAATTTCTGTGAGTCTCTTGCAAAATTTT	951
Qy		301	AspAlaValIleValHisIlysgIuPheTyrrLeuPheTyrrAsnPheGluTyrrPheLeu	320
b		952	GATGCAGATGATGGCACAAACAGTTCATTTGTTTATATTTTGAATATGGTTTTTA	1011
Qy		321	ProNelysPheProPheIleLysIleThrTyrgluGluIleProLeuProIleArgAsn	340
b		1012	CCATATGAAATTCCTCTTTATTTAAATAACATATGAGAATAATCCCTTTAACTATCAGAAAC	1071
Qy		341	LysThrLeuSerGlyLeu	346
b		1072	AAACACTCTCTGGTTTA	1089
RESULT 6				
LOCUS	BD073413	1038 bp	DNA	linear PAT 27-AUG-2002
DEFINITION	Secreted protein which human chromosome 13 encodes.			
ACCESSION	BD073413			
VERSION	BD073413.1 GI:22619016			
KEYWORDS	JF 2001511345-A/12.			
SOURCE	synthetic construct			
ORGANISM	artificial sequences.			
REFERENCE	1 (bases 1 to 1038)			
AUTHORS	Shepard,P.O. and Gilberton,D.G.			
TITLE	Secreted protein which human chromosome 13 encodes			
JOURNAL	Patent: JP 2001511345-A 12 14-AUG-2001;			
ZYMOMENETICS INC				

COMMENT	OS	Artificial Sequence
PN	JP 2001511345-A/12	
PD	14-AUG-2001	
PF	24-JUL-1998 JP 2000504249	
PR	24-JUL-1997 US 60/053613	
PI	PAUL O SHEPPARD, DIBRA G GILBERTON	
PC	C12N15/09, A61K38/00, A61K48/00, C07K14/47, C07K16/18, C12N1/15, PC12N1/19,	
PC	C12N1/21, C12N5/10, C12P21/02, C12Q1/68, C12N15/00, A61K37/02, C12N5/00	
CC	Degenerate nucleotide sequence encoding zsig46 polypeptide of	
CC	SEQ ID NO:2	
FH	Location/Qualifiers	
FT	source	
FT	1..1038	
FT	/organism='Artificial Sequence'	
FT	Location/Qualifiers	
FT	1..1038	
FT	/organism='synthetic construct'	
FT	/mol_type='genomic DNA'	
FT	/db_xref='taxon:32630'	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.05e-139	Length: 1038
Score:	1631.00	Matches: 287
Percent Similarity:	83.19%	Conservative: 0
Best Local Similarity:	83.19%	Mismatches: 58
Query Match:	84.64%	Indels: 0
DB:	6	Gaps: 0
US-10-010-050A-2 (1-346) x BD073413 (1-1038)		
QY	1 MetA r g a r g l y l a g l y l a l a l a r g g l y a r g a l a s e r t r p c y s t r p l a l e u l a l e u 20	
DB	1 A T G M G M G G N G C G N G C G N G C N M G N G G M G N G C N W S N T G T G Y G G C G C N Y T N G C N Y T N 60	
QY	21 L e u t r p l e u l a l a v a l v a l p r o g l y t r p s e r a r g v a l s e r g l y l l e p r o s e r a r g a c h i s 40	
DB	61 Y T N N G Y T N G C G N G T N C C N G N T G G S N M G N G T M S N G S N A T H C C N N S M G M G N C A Y 120	
QY	41 T r p P r o v a l P r o t y r F l y a s r P h e a s p P h e a r g P r o l y s P r o a s p P r o t y r C y s G i n a l a 60	
DB	121 T G G C G C G N G C C N T A A A A M G N T T G A Y T T Y M G N C C N A A C C N G A Y C C N T A A Y T G Y C A R G C N 180	
QY	61 L y s t y l t r p h e c y s p r o t h e l g l y s e r P r o l l e p r o v a l m e t g l u g l y a s p a s p a l l e 80	
DB	181 A A R T A Y A C N T T Y T G C N A C N G M S N C C N A T H C C N G T A T G A R G N A Y G A Y G A Y A T H 240	
QY	81 G l u v a l P h e a r g L e u g l i n a l a p r o v a l T r p g l u p h e l y s t y r g l y a s p l e u l e u g l y h i s 100	
DB	241 G A R G N T T Y M G N Y T M C A R G C N C C N G T N O G G A R T T A A R T A I G G N A Y T T N T N G C N A Y 300	
QY	101 L e u l y s i l e m e t h i a s p a l a l l e g l y P h e a r g s e r t h r L e u t h r g l y l y s a n t y t r t h r 120	
DB	301 Y T N A A R A T H A T C A G A Y G C N A T H G A N T T Y M G N S N A C Y T T A N C N G N A A R A A Y T T A A C N 360	
QY	121 M e t g l u t r p t y r g l u l e u P h e g l i n e u g l y a n c y s t h r P h e P r o h i s t e u a r g P r o g l u 140	
DB	361 A T G G A R T G T A G A Y T N T T Y C A R Y T N G S N A A Y T G A C N T T Y C C N C A Y T T M G N C C N G A R 420	
QY	141 M e t A s p a l a P r o P h e t r p C y s a n g i n g l y a l a a l a c y s P h e p h e g l u g l y l e a s p a s p 160	
DB	421 A T G G A Y G C C C N T T T G G T G Y A Y C A R G G N G C N G C N T G T T Y T T Y G A R G N A T H G A Y G A Y 480	
QY	161 V a l h i s t r p l y g l u a n g l y t h r L e u v a l g l n v a l a t h r l l e s e r g l y a s n m e t P h e 180	
DB	481 G T N C A Y T G A A A G A R A A Y G N A C N Y T N G T N C A R G T N G C N A C N A T H S N G S N A A Y A G T T Y 540	
QY	181 A s n g l i m e t a l e t y t r p v a l l y s g l i n a s p a s n g l u t h g l y l l e t y r t y r g l u t h r t r p 200	
DB	541 A A Y C A R A T G C N A A R T G G T N A R C A R G A Y A Y G A R A C N G N A T T A Y A Y G A R A C N T G G 600	

Accession	Gene	Species	Protein	Length	Size	Score	E-value	RefSeq
201	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
601	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
221	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
661	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
241	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
721	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
261	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
781	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
281	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
841	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
301	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
901	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
321	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
961	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
341	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy
1021	lysPhy	Human	lysPhy	220	220	100.0	0.0	lysPhy

REMARK
COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590
USA

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk

Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),

Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@hgrl.nih.gov
Akhter N. Avelle, K. Beckstrom-

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.T., Granite, S., Guan, Y., Gupta, J., Hachichi, P.

Dieltz, N. L., Grunke, D., Guan, X., Gupta, C., Hagstad, L., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R., Madure, C., Maciel, J. C., Mackay, D., Mastrian, S. D., McLooney, T.

Mauro, V. L., Masriello, C., Masriello, B., Masriello, S. D., McCloskey, J. C., McDowell, J., Pearson, R., Stantirpop, S., Thomas, P. D., Touchman, J. W., Townsend, C., West, T., Weller, M., Wetherill, K. D., Wizinger, J.

Isulgeon, C., Vogt, C. L., Walker, M. A.,
Young, A., Zhang, L.-H. and Green, E. D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL
Series: IRAK Plate: 56 Row: c Column: 7

passed the following selection criteria: Hexamer frequency ORF

analysis. Location/Qualifiers

FEATURES
Source

Location/Qualifiers
1. .2318

```
/organism="Mus musculus"
/mol_type="mRNA"
```

```
/strain="C57BL/6J"  
/db_xref="taxon:10090"
```

```
/clone="IMAGE:5251891"  
/tissue type="Mammary tumor. WAP-TGF alpha model. 7 months
```

old, gross tissue." /clone lib="NCI CGAP Mat

```
/lab_host="DH10B"  
/note="Vector: pCMV-SPORT6"
```

```
<1. .2318
/gene="CTP5"
```

```
/db_xref="LocusID:211286"
/db_xref="MGI:2442253"
```

```
/usr_x86_64-linux-gnu:2142200
<1.1.1013
/canon="C1"nF"
```

```

/ gene = cns
/ codon_start = 3
/ start_pos = 101

```

```

/product="c1ns protein"
/protein_id="AAH25487.1"

```

```

/db_xref="GI:19343/91"
/db_xref="LocusID:211286"

```

```

/db_xref="MGI:2442253"
/translation="TRPAHWRPALALALGLATIGASPTSGQRPVPVYKRFSFRPKT

```

TLTGKNTYTIEMYLEFQIGNCTFPHLRPDKSAFWCNGAACFEGIDDKHWKENGTLSS

KLAEFTIEFKIETNYTKIFLYSGEPIYLGNETSIFGPKGNKTALAIKFFYGPFRPYDNYVWATISGNTFNKVAEMVKQDNETGIIYETWTVRAGPGQGAQTWBBSYDCSNFVLRTYH

DL" LSTKDLELNMFLKIDTVILHRQFYLFYNFEEYWFLPMKBPFEVKITYEETPLPTRHTTFT

8.63e-123 Length: 2318

1451.00	Matches:	258
85.76%	Conservative:	31

key:	76.56%	Mismatches:	42
	75.30%	Indels:	6

10	Gaps :	2
----	--------	---

346) X BC025487 (1-2318)

CTTP-----CysTPAlaLeuAlaLeuLeuTrpLeuAlaValValProGlyTrp 29
..

[illegible]

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GQLY

Center clone name: CH230-331D3

Summary Statistics

Assembly program: Phrap, version 0.990329

Consensus quality: 177481 bases at least Q40

Consensus quality: 179640 bases at least Q30

Consensus quality: 181141 bases at least Q20

Estimated insert size: 183954; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 198524: contig of 198524 bp in length.

Location/Qualifiers

1. 198524

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-331D3"

misc_feature

1. 1183

/note="wgs end extension"

clone end: T7"

131405..131443

/note="clone boundary"

misc_feature

197370..198217

/note="clone boundary"

clone end: Sp6

site: Mbol

end_sequence: RXAPC14TV"

IGIN

Alignment Scores:

ed. No.:

Percent Similarity:

Net Local Similarity:

Very Match:

1:

1-10-010-050A-2 (1-346) x AC109554 (1-198524)

10

38632

GATGAGGCACTGCGCCAGGCGCTCGCGCTAGCTAGCGGCTGCGGCGCACTGTG 38573

10

38632

GATGAGGCACTGCGCCAGGCGCTCGCGCTAGCTAGCGGCTGCGGCGCACTGTG 38573

10

38632

GATGAGGCACTGCGCCAGGCGCTCGCGCTAGCTAGCGGCTGCGGCGCACTGTG 38573

QY 28 GlyTTPSerArgValSerGlyIleProSerArgHISSTPProValProTyrIleArg 47
 Db 38572 GGCCTGCTCCCGACGCTTGGG-----AACGCTGCGCGTACCTTACAGGCGC 38525
 QY 48 PheASPheArgProLysProASPProTyrCysGlnAlaIleTyrThrPheCysProThr 67
 Db 38554 TTTCCTTCCTCCGAGACAGATCCCTCTGTCTGACGCAAGTATCTTCTGCTTACT 38465
 QY 68 GlySerProIleProValMetGluGlyAspAspIleGluValPheArgGluAla 87
 Db 38464 GGCACACCATCCAGATATAGAGACATGACATGACATGACATGACATGACATGAC 38405
 QY 88 ProValTTPGluPheLysTyrGlyAspLeuGluHISLeuValIleMetHISAspAla 107
 Db 38404 CCTTTGGGGAATTTAAATAGAGACCTCTGAGACCTTCAAAATATGACATGACAT 38345
 QY 108 IleGlyPheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGluPhe 127
 Db 38344 ATTGGATTACAGATAGCTTACCTGACCGCAAGACTACATGATGATGATGATGAT 38285
 QY 128 GlnLeuGlyAsnCySerThrPheProHISLeuArgProGluMetAspAlaProPheTyrCys 147
 Db 38284 CAGCTCGGCACTGATACATTTCCCACTTCCGCGGAGTGAAGCTCCGTTCTGATGT 38225
 QY 148 AsnGlnGlyAlaIleCysPhePheGluGlyIleAspValHISThrPheGluAsnGly 167
 Db 38224 AACCAAGGCGGAGCGCTCTCTTGAAGAACTTACGATTAACCTGGAAGGAAACCGG 38165
 QY 168 ThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrVal 187
 Db 38164 ACGTTGCTCTGCTTGGACCATATCCGGAACACGTTTAAACAAAGTGGCGAGTGG 38105
 QY 188 LysGlnAspAsnGluThrGlyIleTyrTyrGluThrPheAsnValLysLysSerProGlu 207
 Db 38104 AACGAGCAATGAGTACCTGGATTTATTAACAGACATGACAGTCCGCGCACCCAGCA 38045
 QY 208 LysGlnAlaGluThrTTPPheAspSerTyrAspCysSerLysPheValLeuArgThrPhe 227
 Db 38044 AAAGGCGCGAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 37985
 QY 228 AsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIle 247
 Db 37984 GAGAAATTTGCTGAATTTGGACAGATTCAGAAAGATGAAAGATGAAAGATGAA 37925
 QY 248 PheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThr 267
 Db 37924 TTCTTTAAGCGGAGAACGATTTACTTGGAAATGAAAGTCTATTTTGGGCCCAA 37865
 QY 268 GlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHISLeu 287
 Db 37864 GGAACCAAGACCTTGTGCTTGGCCATATAAAATTTTACGCGCCCGTCAAAAGGATTCG 37805
 QY 288 ProThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHISLys 307
 Db 37804 TCACCAAAAGATTTCTGTTGAATTTCTTGAATTTTGGACAGATGATATGACACGA 37745
 QY 308 GlnPheTyrLeuPheTyrAsnPheGluTyrTyrPheLeuProMetLysPheProPheIle 327
 Db 37744 GAGTTTACCTGTTTAACTTGAATGATGATGATGATGATGATGATGATGATG 37685
 QY 328 LysIleLeuThrTyrGluGluIleProLeuProIleArgAsnLysThrLysSerGlyLeu 346
 Db 37684 AAATTAACGTACGAGGAAACCCCTTACCTTACCAACATGACATTTTACGACCTG 37628
 RESULT 9 AC107510 227920 bp DNA linear HTG 13-MAY-2003
 AC107510 Rattus norvegicus clone CH230-138J6, *** SEQUENCING IN PROGRESS
 AC107510 *** 4 unordered pieces.
 AC107510.5 GI:30580095
 HTG: HTGS PHASE1, HTGS DRAFT, HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

1 (bases 1 to 227920)
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschrocks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Baidarainke, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cre, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Garroll, L., De Anda, C., Dedrich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flaggs, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisel, A., Ganta, R., Garcia, A., Garner, T., Garzara, W., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulys, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpethy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenswewa, L., Louisedge, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minda, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muihlhass, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Ngunjiri, N., Norris, S., Nwankwelu, O., Okunolu, G., Olarnungsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, B., Pu, L., L., Piazzi, M., Quito, U., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shattman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Soza, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhou, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE

Direct Submission

JOURNAL

2 (bases 1 to 227920)

AUTHORS

Mortley, K. C.

REFERENCE

Submitted (23-JAN-2002)

JOURNAL

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

AUTHORS

Direct Submission

JOURNAL

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 13, 2003 this sequence version replaced gi:22855855. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GJB

Center clone name: CH230-13806

Summary Statistics

Assembly program: Atlas 3.0

Consensus quality: 214528 bases at least Q40

Consensus quality: 217333 bases at least Q30

Consensus quality: 219064 bases at least Q20

Estimated insert size: 223690; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 4 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 223559: contig of 223559 bp in length

223560: gap of unknown length

224739: contig of 1079 bp in length

224839: gap of unknown length

226289: contig of 1451 bp in length

226390: gap of unknown length

226390: contig of 1531 bp in length.

FEATURES

source

1. 227920

misc_feature

/organism="Rattus norvegicus"

misc_feature

/db_xref="taxon:10116"

misc_feature

/clone="CH230-13806"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

misc_feature

/note="wgs contig"

```

37360 TTCTCCTTCCTCGGAGACAGATCCCTTCTGTCAAGCCAAATGATCTTCTGCCCTACT 37419
      68 GlysProileProValMetGluValAspAspAspIleGluValPheArgLeuGlnAla 87
37420 GGCTACACCACTCCAGTATTGAGAGACATGACCTCATCGAAGTCTTAAGGCTACAAACC 37479
      88 ProValTTPGluPheLeuTyrGlyAspLeuLeuGlyHisLeuValIleMetHisAspAla 107
37480 CCGTTTGGAAATTAATATATGAGACCTCTCGGAGACATTCAAATATCATGACAGATCC 37539
      108 IleGlyPheArgSerThrLeuThrGlyValAsnTyrThrMetGluTTPYrGluLeuPhe 127
37540 ATTGATTCAGAGAGACCTGTGACCGGAGAACTACACATTAATGATGATGAGCTTTTC 37599
      128 GlnLeuGlyAsnCyThrPheProHisLeuArgProGluMetAspAlaProPheTyrCys 147
37600 CAGCTCGGCAACTGATACATTTCCCACTCGGCGGAGAGAAAGCGTCCGTCGTGGTGT 37659
      148 AsnGlnGlyAlaAlaAcYpPhePheGluGlyIleAspAspValHisTyrPlyGluAsnGly 167
37660 AACCGAGGGGCGAGCCTCTCTTTGAGAGAAATTTACATTAACACTCGAGAGAAACGGG 37719
      168 ThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGluMetAlaIleTyrVal 187
37720 ACGTTGTCTCTGGTGGCAACCATTCGGAACACGTTTAACAAATGAGCGAGTGGGTG 37779
      188 LysGlnAspAsnGluThrGlyIleTyrTyrGluThrTTPAsnValIleAlaSerProGlu 207
37780 AAGCAGCAATAGACACTCGGATTTATTACGAGCATGAGACGTCCGGGCGGACCCAGGA 37839
      208 LysGlyAlaGluThrTTPheAspSerTyrAspCysSerLysPheValLeuArgThrPhe 227
37840 AAGGCGGCGGAGACATGTTTGAATCTTGAACCTGTGCAATTTGTCTTAAGGACAAAT 37899
      228 AsnValLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIle 247
37900 GAGAAATTTGGCTGAAATTTGGAACAGATTCAGAAAGATGAGAAACAACTATATGAAATA 37959
      248 PheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThr 267
37960 TTCTTTTACAGGGGAGAAACCGATTTACTTGGGAAATGAAACGTCTATTTTGGGCCAAA 38019
      268 GlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeu 287
38020 GGAACCAAGACCTTGTGCTTGGCCATTAATAAATTTTACGGCCCGTCMAAGGATTCG 38079
      288 ProThrLysGluPheLeuLeuSerLeuGlnIlePheAspAlaValIleValHisLys 307
38080 TCACCAAGAGATTTCTGTGTAATTTCTGAAATTTTGGACACAGTGAATTATGACAGA 38139
      308 GlnPheTyrLeuPheTyrAsnPheGluTyrTyrPheLeuProMetLysPheProPheIle 327
38140 GAGTCTCACTGTTTATTAACCTTGAGTATGCTTCTTCACTATGAAACCCCTTTGTGC 38199
      38140 GAGTCTCACTGTTTATTAACCTTGAGTATGCTTCTTCACTATGAAACCCCTTTGTGC 38199
SUBUT 10
1125187 697 bp DNA linear PART 18-SEP-2002
3125187
3125187 Primer for synthesizing full-length cDNA and use thereof.
3125187 BD125187.1 GI:23220132
3125187 UP 2002017375-A/618.
3125187 31WORDS
3125187 JRCRC
3125187 Homo sapiens (human)
3125187 ORGANISM
3125187 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
3125187 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
3125187 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
3125187 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
3125187 Koga,H.
3125187 Primer for synthesizing full-length cDNA and use thereof
3125187 JOURNAL Patent: JP 2002017375-A 618 22-JAN-2002;
3125187 HELIX RESEARCH INSTITUTE
3125187 OS Homo sapiens (human)

```

```

PN JP 2002017375-A/618
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUO OTSUKI, HISASHI KOGA
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/
10, C12P21/02, C12P1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
PC Primer for synthesizing full-length cDNA and use thereof FH Key
FT source Location/Qualifiers
FT 1..697
FT location/Qualifiers
FT 1..697
FT /organism="Homo sapiens"
FT /molecule="genomic DNA"
FT /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 3.7e-88 Length: 697
Score: 1065.00 Matches: 200
Percent Similarity: 93.55% Conserved: 3
Best Local Similarity: 92.17% Mismatches: 8
Query Match: 55.27% Indels: 6
DB: Gaps: 2
US-10-010-050A-2 (1-346) x BD125187 (1-697)
QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTyrCysTTPAlaLeu 20
DB 52 ATGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 111
QY 21 LeuTTPLeuAlaValAlaProGlyTTPSerArgValSerGlyIleProSerArgHis 40
DB 112 CTGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 171
QY 41 TTPProValProTyrLysArgPheAspPheArgProLysProAspProTyrCysGlnAla 60
DB 172 TGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 231
QY 61 LysTyrThrPheCysProThrGlySerProIleProValMetGluGlyAspAspAspIle 80
DB 232 AAGTATACCTTCTGTGCACTGCTCACCTATCCAGTTATGAGGAGATGATGACACT 291
QY 81 GluValPheArgLeuGlnAlaProValTTPGluPheLysTyrGlyAspLeuLeuHis 100
DB 292 GAACTTTTCAATTACAGAGCCCGCATGAGAAATTAATATGAGACCTCTGGGACAC 351
QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
DB 352 TTGAAATTTATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411
QY 121 MetGluTTPYrGluLeuPheGlnLeuGlyAsnCyThrPheProHisLeuArgProGlu 140
DB 412 ATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 471
QY 141 MetAspAlaProPheTTPCysAsnGlnIleGlyAlaAcYpPhePheGluGlyIleAspAsp 160
DB 472 ATGAGTGGCCCTTCTGTGATATGAGGCGCTGCTCTTTTGGAGGATTTATGAT 531
QY 161 ValHisTTPYrGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetLys 180
DB 532 GTTCACTGGAAGGAGAAAGGAGCACTTAAGTCAATAGCAATATATGAGGAGAAACATGT 591
QY 180 heAsnGlnMetAla-LysTyrValLysGln--AspAsnGluThrGlyIleTyrTyrGlu 198
DB 592 TCACCAATATGAGCAATGAGGAGTGAACAGAGCAATGAGCAATGAGCAATGAGCAATGAG 651

```

199 -ThrtPrsAnVAllysaLaserProgluLys---GlyAlaGlu 211
 652 AACATGGAATTTTAAAGCCANCCANNAANAGGGGCAAAA 694

RESULT 11
 ID126453
 ACUS BD126453 697 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD126453.1 GI:23221398
 VERSION BD126453.1
 KEYWORDS JP 2002017375-A/1884.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 697)
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/1884
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO I, ISHII,
 PI YUKI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI SHINICHI KOJIMA,
 PI TETSUO OTSUKI, HISASHI KOGA
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..697
 Location/Qualifiers
 1..697
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores:
 Pred. No.: 3.7e-88 Length: 697
 Score: 1065.00 Matches: 200
 Percent Similarity: 93.55% Conservative: 3
 Best Local Similarity: 92.17% Mismatches: 6
 Query Match: 55.27% Indels: 2
 Gaps: 2

US-10-010-050a-2 (1-346) x BD126453 (1-697)

QY 1 MetarhArgGlyAlaGlyAlaAlaArgGlyAlaGlyAlaSerTrpCysTrpAlaLeuAlaLeu 20
 DB 52 ATGGCGGG 111
 QY 21 LeuTrpLeuAlaAlaAlaAlaProGlyTrpSerArgValSerGlyIleProSerArgArgHis 4C
 DB 112 CTTTGGGCTCGGG 171
 QY 41 TrpProValProTrpValArgPhaAspPheArgProLysProAspProTrpCysGlnAla 6C
 DB 172 TGGCGGG 231
 QY 61 LysTrpThrPheCysProThrGlySerProIleProValMetGluGlyAspAspAspIle 80
 DB 232 AAGTAATCTTCTGTCTCACTGGCTCACTATCCAGTTATGAGAGGGGATGAGATGACACT 291
 QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTrpGlyAspLeuLeuGlyHis 100

DB 292 GAACTTTTGCATACAGAGCCCGAGTGGATTTAAATGAGACCTTCCTGGACAC 351
 QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuTrpGlyLysAsnTrpThr 120
 DB 352 TTGAAATATATGACATGATGCCATTGGATTCCAGAGTACATTAACTGGCANAACTACACA 411
 QY 121 MetGluTrpTrpGluLeuPheGlnLeuGlyAsnGlyTrpPheProHisLeuArgProGlu 140
 DB 412 ATGGAATGATGAGAACTTTTCAACTGGCAACTGTACATTCCCATCTCCGACCTGAA 471
 QY 141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGlnGlyIleAspAsp 160
 DB 472 ATGATGCCCCCTTCTGTGTATCAAGCGCGCTGCTTTTGGAGGAATTTATGAT 531
 QY 161 ValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetCp 180
 DB 532 GTTCACTGGAAGGAAATATGGACATTAGTTCAGTACCACTATATCAGGAAACATAGT 591
 QY 180 heAsnGlnMetAla-LysTrpValIleGln---AspAsnGluThrGlyIleLysTrpGlu 198
 DB 592 TCAACCAATGGCAAAAGTGGGTGAACAGACAACTGAAACAGGAATTTATTATGAG 651

199 -ThrtPrsAnVAllysaLaserProgluLys---GlyAlaGlu 211
 652 AACATGGAATTTTAAAGCCANCCANNAANAGGGGCAAAA 694

RESULT 12
 AC001226 106988 bp DNA linear PRI 29-MAY-1997
 LOCUS Genomic sequence from Human 13, complete sequence.
 DEFINITION AC001226
 ACCESSION AC001226.1 GI:2133862
 VERSION AC001226.1
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 106988)
 Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W., Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P., Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Frittp,W.J., Gage,D., Gerliger,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J., Margulis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J., Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and Zody,M.
 Genomic sequence from Human 13
 Unpublished
 2 (bases 1 to 106988)
 Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W., Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P., Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Frittp,W.J., Gage,D., Gerliger,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J., Margulis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J., Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and Zody,M.
 Direct Submission
 Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 106988)
 Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W., Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Frittp,W.J., Gage,D., Gerliger,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J., Margulis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J., Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and Zody,M.
 Direct Submission
 Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 29, 1997 this sequence version replaced as 1932720.
 The staden databases, finishing information, and all chromatographic files used in the assembly of this clone are available from our anonymous ftp site.

All repeats were identified using RepeatMasker: Smit, A.F.A. &

Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
Location/Qualifiers

FEATURES
source

```
1.106988
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="76N15"
/clone_1ib="unknown"
complement(113..200)
/rpt_family="TRE1B"
complement(202..611)
/rpt_family="L1PA3"
complement(674..694)
/rpt_family="AT_rich"
complement(835..981)
/rpt_family="MIR2"
1632..1920
/rpt_family="AluJo"
1970..2271
/rpt_family="AluSp"
2779..2946
/rpt_family="FAM"
3552..3828
/rpt_family="AluJb"
4361..4640
/rpt_family="AluSg"
5253..5385
/rpt_family="MLTIF"
5914..6106
/rpt_family="MLTIF"
complement(6341..6435)
/rpt_family="MER34"
complement(6838..7140)
/rpt_family="AluSx"
7394..7455
/rpt_family="MERSA"
7464..7627
/rpt_family="MERSB"
complement(7624..7792)
/rpt_family="MERSB"
complement(8363..8391)
/rpt_family="AT_rich"
8480..8639
/rpt_family="MERSB"
complement(9199..9286)
/rpt_family="(TA)n"
9340..9644
/rpt_family="AluJo"
9880..9946
/rpt_family="MIR2"
10208..10787
/rpt_family="L1MB3"
10984..11297
/rpt_family="AluSg"
11311..11470
/rpt_family="L1PA8"
complement(11471..11495)
/rpt_family="AT_rich"
complement(12394..12440)
/rpt_family="AT_rich"
13782..13921
/rpt_family="MIR"
14035..14316
/rpt_family="AluSx"
complement(15018..15046)
/rpt_family="AT_rich"
15080..15373
/rpt_family="AluJb"
15588..15881
/rpt_family="AluJo"
complement(18038..18071)

/rpt_family="(CAT)n"
complement(20389..20494)
/rpt_family="FAM"
21238..21534
/rpt_family="AluSx"
complement(23331..23353)
/rpt_family="GC_rich"
complement(23546..23578)
/rpt_family="GC_rich"
24565..24707
/rpt_family="MIR"
26013..26047
/rpt_family="MIR2"
complement(27717..28532)
/rpt_family="L1PB1"
complement(28553..28854)
/rpt_family="AluSg"
complement(28860..29076)
/rpt_family="L1"
complement(29365..29561)
/rpt_family="MIR"
complement(30050..30147)
/rpt_family="MIR"
31487..31812
/rpt_family="AluJo"
complement(33629..33732)
/rpt_family="MIR"
34434..34729
/rpt_family="AluSx"
35070..35361
/rpt_family="AluSx"
complement(36437..36461)
/rpt_family="AT_rich"
complement(39967..40261)
/rpt_family="AluSx"
complement(40741..40915)
/rpt_family="AluSg"
40963..41663
/rpt_family="L1PA2"
42042..42352
/rpt_family="MLTIF"
42467..42771
/rpt_family="AluSg"
42772..42848
/rpt_family="L1MB8"
42849..43151
/rpt_family="AluJb"
43152..43318
/rpt_family="L1MA10"
43253..43427
/rpt_family="L1MB8"
43429..43672
/rpt_family="L1MB8"
complement(44188..44481)
/rpt_family="AluSc"
44544..45246
/rpt_family="L1TR8"
complement(46999..47032)
/rpt_family="AT_rich"
complement(47411..47431)
/rpt_family="AT_rich"
complement(47612..47684)
/rpt_family="(CAT)n"
complement(47946..48239)
/rpt_family="AluJo"
complement(49827..49872)
/rpt_family="AT_rich"
complement(50160..50186)
/rpt_family="AT_rich"
50952..51121
/rpt_family="MERSA"
complement(52140..52284)
/rpt_family="AluJo"
```

```

repeat_region      53772..53974
repeat_region      54225..54348
repeat_region      complement(54533..54579)
repeat_region      /rpt_family="MIR"
repeat_region      complement(54580..54736)
repeat_region      /rpt_family="AluJb"
repeat_region      complement(54755..54782)
repeat_region      /rpt_family="(CA)n"

Alignment Scores:
pred. No.:          1,41e-72      Length:      106988
score:              925.50        Matches:      173
Percent Similarity: 89.69%        Conservative: 1
Best Local Similarity: 89.18%      Mismatches: 1
Query Match:        48.03%         Indels:      19
JB:                 9             Gaps:        1

JS-10-010-050a-2 (1-346) x AC001226 (1-106988)

2Y      153  CysPhePheGluGlyLeuAspValHisTrpLysGluAsnGlyThrLeuValGlnVal 172
Db      31851 TGTGTTTTTAAA----- 31362

2Y      173  AleThrIleSerGlyAsnMetPheAsnGlnMetAlaYserTrpValLysGlnAspAsnGln 192
Db      31863 -----CTAGGAAACATGTTTCAACCAATGGCAAAAGGGGTGAAACAGCAATGAA 31913

2Y      193  ThrGlyIleTyTrpGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGluThr 212
Db      31914 ACAGGAATTTATATATGAGCATGGAATGTAAAGCCAGCAAGAAAGGGGGCAGAGACA 31973

2Y      213  TrpPheAspSerTyTrpAspCysSerLysPheValLeuArgThrPheAsnLysLeuAlaGlu 222
Db      31974 TGGTTTATTCCTTACGACCTGTTCCAAATTTGTGTTAAGACCTTTTAAAGTTGGCTGAA 32033

2Y      233  PheGlyAlaGluPheLysAsnIleGluThrAsnTyThrArgIlePheLeuTySerGly 252
Db      32034 TTTCGAGCAGAGTTCAACCAATGAAACCACTATTCAGAAATTTCTTTACAGTGA 32093

2Y      253  GluProThrTyLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeu 272
Db      32094 GAACCTACTTATCTGGGAAATGAACAATCTGTTTGGGCCAACAGAAACAAGACTCTT 32153

2Y      273  GlyLeuAlaIleLysArgPheTyTrpProPheLysProHisLeuProThrLysGluPhe 292
Db      32154 GGTTAGCCATAAAGATTTTATTACCCCTTCAACCACTTTGCCAAGCTTAAAGAAATTT 32213

2Y      293  LeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyLeuPhe 312
Db      32214 CTGTTGAGTCTCTGCAAAATTTTGAAGCAGTATTTGTGCACAAACAGTTCTATTGTTT 32273

2Y      313  TyAsnPheGluTyTrpPheLeuProMetLysPheProPheIleLysIleThrTyGlu 332
Db      32274 TATATATTTGAAATTTGTTTATTACCTATGAAATTCCTTTTATTAATAATCATATGAA 32333

2Y      333  GluIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 346
Db      32334 GAAATCCCTTACCTATCATGAAACAAACACTCTCTGTGTTA 32375

RESULT 13
AL136440      169362 bp      DNA      linear      HTG 10-JUN-2001
LOCUS      Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.
DEFINITION      AL136440
ACCESSION      AL136440.2 GI:10039473
VERSION      HTG; HTGS PHASE1; HTGS_CANCELLED.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1

```

AUTHORS TITLE JOURNAL COMMENT

Burton, J.
Direct Submission
Submitted (10-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Sep 8, 2000 this sequence version replaced gi:6982057.
Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
Project Information
Center project name: ba185124
Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 162085 bases at least Q40
Consensus quality: 164521 bases at least Q30
Consensus quality: 166040 bases at least Q20
Insert size: 168062; sum-of-contigs
Insert size: 164996; 1.7% error; agarose-fp
Quality coverage: 4.08x in Q20 bases; sum-of-contigs Quality
Coverage: 4.16x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      3400: contig of 3400 bp in length
*      3401      3500: gap of 100 bp
*      3501      16444: contig of 12344 bp in length
*      16445      16544: gap of 100 bp
*      16545      19514: contig of 2970 bp in length
*      19515      22241: contig of 2627 bp in length
*      22242      22341: gap of 100 bp
*      22342      25526: contig of 3155 bp in length
*      25527      25626: gap of 100 bp
*      25627      33070: contig of 7444 bp in length
*      33071      33170: gap of 100 bp
*      33171      67620: contig of 34450 bp in length
*      67621      67720: gap of 100 bp
*      67721      93413: contig of 25693 bp in length
*      93414      93513: gap of 100 bp
*      93514      114081: contig of 20868 bp in length
*      114082      114181: gap of 100 bp
*      114182      126206: contig of 12025 bp in length
*      126207      126306: gap of 100 bp
*      126307      144583: contig of 18277 bp in length
*      144584      144683: gap of 100 bp
*      144684      147136: contig of 2453 bp in length
*      147137      147236: gap of 100 bp
*      147237      164925: contig of 17689 bp in length
*      164926      165025: gap of 100 bp
*      165026      169362: contig of 4337 bp in length.
Location/Qualifiers
1. 169362
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-185124"
/clone_1bp="RPCT-11.1"
1. 3400
/note="assembly_fragment:00318
clone_end:17
vector_side:left"
3501..16444

```

FEATURES

source
misc_feature
misc_feature
misc_feature

*	197683	197782:	gap of 100 bp	
*	197783	201788:	contig of 4006 bp	in length
*	201789	201888:	gap of 100 bp	in length
*	201889	204060:	contig of 2172 bp	in length
*	204061	204160:	gap of 100 bp	in length
*	204161	206237:	contig of 2077 bp	in length
*	206238	206337:	gap of 100 bp	in length
*	206338	210376:	contig of 4039 bp	in length
*	210377	210476:	gap of 100 bp	in length
*	210477	225245:	contig of 14769 bp	in length
*	225246	225345:	gap of 100 bp	in length
*	225346	230027:	contig of 4682 bp	in length
*	230028	230127:	gap of 100 bp	in length
*	230128	233610:	contig of 2483 bp	in length
*	232611	232710:	gap of 100 bp	in length
*	232711	234942:	contig of 2232 bp	in length
*	234943	235042:	gap of 100 bp	in length
*	235043	240958:	contig of 5516 bp	in length
*	240959	241058:	gap of 100 bp	in length
*	241059	251187:	contig of 10129 bp	in length
Location/Qualifiers				
1. .251187				
/organism="Homo sapiens"				
/mol_type="genomic DNA"				
/db_xref="taxon:9606"				
/chromosome="13"				
/clone="RP11-453N22"				
/clone_id="RPIC1-11.2"				
1. .2130				
misc_feature	/note="assembly fragment: 00469			
	fragment chain:1			
	2231. .5689			
misc_feature	/note="assembly fragment: 01219			
	fragment chain:1			
	5790. .10046			
misc_feature	/note="assembly fragment: 00352			
	fragment chain:1			
	10147. .12624			
misc_feature	/note="assembly fragment: 00547			
	fragment chain:2			
	12725. .20993			
misc_feature	/note="assembly fragment: 00617			
	fragment chain:2			
	21094. .28936			
misc_feature	/note="assembly fragment: 01248			
	fragment chain:2			
	29037. .32230			
misc_feature	/note="assembly fragment: 00606			
	fragment chain:3			
	32331. .35153			
misc_feature	/note="assembly fragment: 01016			
	fragment chain:3			
	35254. .37649			
misc_feature	/note="assembly fragment: 01277			
	fragment chain:3			
	37750. .40245			
misc_feature	/note="assembly fragment: 00920			
	fragment chain:4			
	40346. .48420			
misc_feature	/note="assembly fragment: 01288			
	fragment chain:4			
	48521. .52169			
misc_feature	/note="assembly fragment: 00615			
	fragment chain:4			
	52270. .56443			
misc_feature	/note="assembly fragment: 00009			
	fragment chain:5			
	56544. .63068			
misc_feature	/note="assembly fragment: 00448			
	fragment chain:5			
	63189. .65281			
misc_feature	/note="assembly fragment: 00027			

TITLE Functional and comparative genomic analysis of the piebald deletion

Matches:	127
Conservative:	22
Similarity:	87.65%
	730.00

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

un on: April 24, 2004, 22:59:32 ; Search time 5973.28 Seconds
(without alignments)
2292.945 Million cell updates/sec

file: US-10-010-050A-2_COPY_31_346
erfect score: 1758
equence: 1 RVSGTSPRRKMPVYKRFDF.....IKTYEIRIPINKTUSGL 316

coring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

searched: 3470272 seqs, 21671516995 residues

total number of hits satisfying chosen parameters: 6940544

minimum DB seq length: 0
maximum DB seq length: 2000000000

post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

command line parameters:

MODEL=frame+ p2n model -DEV=xlh
Q=/cgn2.1/USPTO.epool/US10010050/runat.22042004.113204.27549/app.query.fasta_1.1372
DB=genembl -QWMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi -LIST=45
DOCLINK=200 -THR.SCOR=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
OUTMT=ptco -NORM=ext -HEPFSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US10010050 @cgn.1.17433 @runat.22042004.113204.27549 -NCPU=6 -ICPU=3
NO.MMAP -LARGECQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : GenEmbl.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_pi.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ets.*
12: gb_by.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_ov.*
22: em_or.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vtc.*
38: em_by.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match Length	DB ID	Description
1	1758	100.0 1486	BD073402	BD073402 Secreted
2	1758	100.0 4080	AF068227	AF068227 Homo sapi
3	1747	99.4 1751	BD136338	BD136338 95 human
4	1744	99.2 2120	BD127858	BD127858 primer fo
5	1744	99.2 2120	AK075109	AK075109 Homo sapi
6	1515	86.2 1038	BD073413	BD073413 Secreted
7	1423	80.9 2318	BC025487	BC025487 Mus muscu
8	1396	79.4 198524	AC109554	AC109554 Rattus no
9	1339	76.2 227920	AC107510	AC107510 Rattus no
10	925.5	52.6 106988	AC001226	AC001226 Genomic s
11	925.5	52.6 169362	AL136440	AL136440 Homo sapi
12	925.5	51.6 221187	BD1359875	BD1359875 Homo sapi
13	896	51.0 697	BD125187	BD125187 primer fo
14	896	51.0 697	BD126453	BD126453 primer fo
15	730	41.5 110000	AE014175_0	AE014175 Mus muscu
16	730	41.5 156823	AC102815	AC102815 Mus muscu
17	727	40.4 506	BD060465	BD060465 Secreted
18	716	40.7 229015	AC106966	AC106966 Rattus no
19	711	40.4 217336	AC131344	AC131344 Rattus no
20	617.5	35.1 251187	AL359875	AL359875 Homo sapi
21	582	33.1 198524	AC109554	AC109554 Rattus no
22	272.5	15.5 258273	AC113855	AC113855 Rattus no
23	266	14.6 110000	AC116305_0	AC116305 Dictyoste
24	241	13.7 473	AX333881	AX333881 Sequence
25	241	13.7 473	AX335993	AX335993 Sequence
26	146.5	8.3 256879	AC116982	AC116982 Dictyoste
27	111.5	6.3 223182	AL929049	AL929049 Zebrafish
28	104.5	5.9 2162	BC014298	BC014298 Homo sapi
29	104.5	5.9 2175	AX881019	AX881019 Sequence
30	104.5	5.9 2175	BD158729	BD158729 primer fo
31	104.5	5.9 2175	AK023964	AK023964 Homo sapi
32	103	5.9 1634	BD275044	BD275044 50 Human
33	103	5.9 2904	AK075480	AK075480 Homo sapi
34	103	5.9 171109	EX005356	EX005356 Zebrafish
35	101	5.7 1638	AY198311	AY198311 Streptoco
36	101	5.7 1946	AF407190	AF407190 Trichophy
37	100	5.7 55686	AF407186	AF407186 Trichophy
38	100	5.7 110000	AC095291_3	AC095291 Continuation (3 of
39	100	5.7 110000	AC095291_2	AC095291 Continuation (4 of
40	100	5.7 165504	AC140352	AC140352 Mus muscu
41	100	5.7 175765	AL683871	AL683871 Human DNA
42	100	5.7 219318	AC133992	AC133992 Rattus no
43	100	5.7 251506	AC106702	AC106702 Rattus no
44	99.5	5.7 305961	AE016937	AE016937 Bacteroid
45	99	5.6 140332	AC118538	AC118538 Felis cat

RESULT 1

ALIGNMENTS

BD073402 1486 bp DNA linear PAT 27-AUG-2002
LOCUS Secreted protein which human chromosome 13 encodes.
DEFINITION BD073402
ACCESSION BD073402.1 GI:22619005
VERSION JP 2001511345-A/1.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1486)
AUTHORS Sheppard,P.O. and Gilberton,D.G.
TITLE Secreted protein which human chromosome 13 encodes
JOURNAL Patent: JP 2001511345-A 1 14-AUG-2001;
ZMOGENETICS INC
COMMENT OS Homo sapiens (human)
PN JP 2001511345-A/1
PD 14-AUG-2001
PF 24-JUL-1998 JP 200504249
PR 24-JUL-1997 US 60/053613
PI PATL O SHEPPARD, DIBRA G GILBERTON
PC C12N15/09,A61K38/00,A61K48/00,C07K14/47,C07K16/18,C12N1/15,
C12N1/19,
PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,C12N15/00,A61K37/02,C12N5/PC
00
CC Secreted protein which human chromosome 13 encodes FH Key
FT CDS Location/Qualifiers
1..1486
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2,95e-170 Length: 1486
Score: 1758.00 Matches: 316
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-010-050a-2_COPY_31_346 (1-316) x BD073402 (1-1486)
QY 1 ArgValSerGlyLeuProSerArgHisTyrProValProTyrLysArgPheAspPhe 20
DB 137 CGGGTCTCGGGGCAATCCCTCCCGCGCCACTGGCCGGTGCCTACAAAGCCCTTGAATTG 196
QY 21 ArgProLysProAspProTyrCysGlnAlaLysTyrPheCysProThrGlySerPro 40
DB 197 CGTCAAAACCTGATCTTATGTCAGCTAGATATCTTCTGCTCCACTGCTACCT 256
QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyr 60
DB 257 ATCCCACTTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 316
QY 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleLeuMetHisAspAlaIleGlyPhe 80
DB 317 GAATTTAAATATGAGACCTCCCTGGGACACTTGAATTTATGATGATGATGATGATGATG 376
QY 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGly 160
DB 377 AGAAGTACTATTAATGCGCAAGACTACCAATGCAATGCAATGCAATGCAATGCAATG 436
QY 101 AsnGlyThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
DB 437 AACTGTACATTTCCCATCTCCGACCTGAATGATGATGATGATGATGATGATGATGATG 456
QY 121 AlAlaLysPhePheGlnGlyIleAspAspValHisTyrLysGlnAsnGlyThrLeuVal 140
DB 497 GCTGCTGCTTTTGTGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 556

QY 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrValLysGlnAsp 160
DB 557 CAAGTAGCACTATATATGAGAAACATGTTCAACCAATGCGAAAGTGGGTAAACAGAC 616
QY 161 AsnGluThrGlyIleTyrTyrGluThrTyrAsnValLysAlaSerProGluLysGlyAla 180
DB 617 AATGAAACGGAATTTATTAAGACATGAAATGAAAGGACGCGCAAGAAAGGGGGA 676
QY 181 GluThrTyrPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200
DB 677 GAGCATATGTTGATCTCTTACGACCTGTTCCAAATTTGCTTAAGACCTTTTACAGATTG 736
QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220
DB 737 GCTGATTTTGGAGAGAGATTCAAGAACATGAAACCACTATACAGATATATTTCTTAC 796
QY 221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
DB 797 AGTGAGAACTACTATCTGGAAATGAAACATCTGTTTGGGCGCAACAGAAACAG 856
QY 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLys 260
DB 857 ACTCTGTTTACGCAATTAAGATTTATTAACCTTCAACCAACCATTTGCCCACTAA 916
QY 261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyr 280
DB 917 GAATTTCTGTTGAGCTCTCTTGAATTTTGTATGATGATGATGATGATGATGATGATG 976
QY 281 LeuPheTyrAsnPheGluTyrTyrPheLeuProMetLysPheProPheIleLysIleThr 300
DB 977 TTGTTTAAATTTGAAATATGTTTCTTACCTATGAAATTTCTTTATTTAAATTAACA 1036
QY 301 TyrGluGluIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 316
DB 1037 TATGAAAGAAATCCCTTACCTATCAGAAACAAACAACTCTGCTTGA 1084
RESULT 2
AF068227 4080 bp mRNA linear PRI 28-UTL-1998
LOCUS AF068227
DEFINITION Homo sapiens putative transmembrane protein (CIN5) mRNA, complete
cde
ACCESSION AF068227
VERSION AF068227.1 GI:3342385
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4080)
AUTHORS Savukoski,M., Klockars,T., Holmberg,V., Santavuori,P., Lander,E.S.
and Peltonen,L.
TITLE CIN5, a novel gene encoding a putative transmembrane protein
mutated in Finnish variant late infantile neuronal ceroid
lipofuscinosis
JOURNAL Nat. Genet. 19 (3), 286-288 (1998)
MEDLINE 98324783
PUBMED 9662406
REFERENCE 2 (bases 1 to 4080)
AUTHORS Savukoski,M., Klockars,T., Holmberg,V., Santavuori,P., Lander,E.S.
and Peltonen,L.
TITLE Direct Submersion
SUBMITTED (26-MAY-1998) Department of Human Molecular Genetics,
National Public Health Institute, Mannerheimintie 166, Helsinki
00300, Finland
JOURNAL Location/Qualifiers
1..4080
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q22"
1..4080
gene

ORIGIN

Alignment Scores:

Pred. No.: 4,886-169 Length: 1751
Score: 1747.00 Matches: 314
Percent Similarity: 99.37% Conservative: 2
Best Local Similarity: 99.37% Mismatches: 0
Query Match: 99.37% Indels: 0
Gaps: 0

US-10-010-050A-2_COPY_31_346 (1-316) x BD136338 (1-1751)

```
2Y 1 ArgValSerGlyIleProSerArgHISerProValProTyrLysArgPheAspPhe 20
Db 139 CGGGTCTCGGGCAATCCCTCCGGCCGCTGCGCGCTGACCAAGGCTTGACTTC 193
2Y 21 ArgProLysProAspProTyrCysGlnAlaIleTyrThrPheCysProThrGlySerPro 40
Db 199 CGTCCAAAACCTGATCTTATGTCTCAAGCTAAGTATATCTTCTGTCCAACTGGCTCACT 258
2Y 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValITP 60
Db 259 ATCCCACTTATGAGGGGTGATGATGACATTGAAGTTTTCATTACAAAGCCCAAGTATGG 318
2Y 61 GluPheLysTyrGlyAspLeuLeuGlyHISLeuIleIleMetHISAspAlaIleGlyPhe 80
Db 319 GAATTTAAATATGAGAGACCTCTGGGACACTTGAAATATGATGATGATGCTTGAATTC 378
2Y 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrGlyLeuPheGlnLeuGly 100
Db 379 AGAGTACATTAACTGCAAGAACTACACAAATGATGATGATGATGATGATGATGATGATG 438
2Y 101 AsnGlySerThrPheProHISLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
Db 439 AACGTACATTTCCTCCCACTCCGACCTGAAATGATGATGATGATGATGATGATGATGAT 458
2Y 121 AlaIleCysPhePheGlnGlyIleAspAspValHISerTyrLysGlnLeuVal 140
Db 499 GCTGCTCTCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 558
2Y 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrValLysGlnAsp 160
Db 559 CAAGTAGCACTATATCAGAAACATGTTCAACCAATGGAAGTGGTGGTGGTGGTGGTGG 618
2Y 161 AsnGluThrGlyIleTyrTyrGluThrTyrAsnValLysAsnSerProGluLysGlyAla 180
Db 619 AATGAAACAGAAATTTATATGACATGATGATGATGATGATGATGATGATGATGATG 678
2Y 181 GluThrTyrPheAspSerTyrAspCysSerLysPheValIleuArgThrPheAsnLysLeu 200
Db 679 GAGCATGTGTTGATCTCTACGACTGTTCCAAATTTGTGTAAAGACCTTTAAACAAGT 738
2Y 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220
Db 729 GGTGAATTTGAGAGCAGAGTTCACAGAACATAGAACCACTATACAFGAATATTTCTTAC 798
2Y 221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
Db 799 AGTGAGAGACCTATCTATCTGGGAATGAAACATCTGTTTGGGCCCAACGAGAAACAG 858
2Y 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHISLeuProThrLys 260
Db 859 ACTCTGTGTTAGCATAAAAGATTTATATACCTTCAAAACCAATTTGGCACTAA 918
2Y 261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHISLysGlnPheTyr 280
Db 919 GAATTTCTGTGAGTCTCTGCAAAATTTTGTATGAGAGATTTGTGACAAACAGAGTTTAT 978
2Y 281 LeuPheTyrAspPheGluTyrTyrPheLeuProMetLysPheProPheIleLysIleThr 300
Db 979 TTGTTTAAATTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1038
2Y 301 TyrGluGluIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 316
```

Db 1039 TATGAAAGAAATCCCTTATCTATACAGAAACAAACACTCTCTGTTTAA 1086

RESULT 4
LOCUS BD127858 2120 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD127858
VERSION BD127858.1 GI:23222803
KEYWORDS JP 2002017375-A/3289.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota,T., Nishikawa,T., Ieogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.

TITLE
JOURNAL Primer for synthesizing full-length cDNA and use thereof
PATENT: JP 2002017375-A 3289 22-JAN-2002;
HELIX RESEARCH INSTITUTE

COMMENT OS Homo sapiens (human)
PN JP 2002017375-A/3289

PD 22-JAN-2002 JP 2002053172
PF 07-JUN-2000 JP 2002053172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOHAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUO OTSUKI, HISASHI KOGA

PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ 10,
10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
PC Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS (16)..(1089).

FEATURES
source 1..2120
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 1,278-168 Length: 2120
Score: 1744.00 Matches: 314
Percent Similarity: 99.37% Conservative: 0
Best Local Similarity: 99.37% Mismatches: 2
Query Match: 99.20% Indels: 0
Gaps: 0

US-10-010-050A-2_COPY_31_346 (1-316) x BD127858 (1-2120)

```
2Y 1 ArgValSerGlyIleProSerArgHISerProValProTyrLysArgPheAspPhe 20
Db 142 CGGGTCTCGGGCAATCCCTCCGGCCGCTGCGCGCTGACCAAGGCTTGACTTC 201
2Y 21 ArgProLysProAspProTyrCysGlnAlaIleTyrThrPheCysProThrGlySerPro 40
Db 202 CGTCCAAAACCTGATCTTATGTCTCAAGCTAAGTATATCTTCTGTCCAACTGGCTCACT 261
2Y 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValITP 60
Db 262 ATCCCACTTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 321
2Y 61 GluPheLysTyrGlyAspLeuLeuGlyHISLeuIleIleMetHISAspAlaIleGlyPhe 80
Db 322 GAATTTAAATATGAGAGACCTCTGGGACACTTGAAATATGATGATGATGATGATGATG 381
2Y 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrGlyLeuPheGlnLeuGly 100
Db 382 AGAGTACATTAACTGCAAGAACTACACAAATGATGATGATGATGATGATGATGATGATG 441
```

101 AsnCystrhpheProHisLeuArgProGlnMetAspAlaProhetPrpCysAsnGlnGly 120
442 AACTGTACATTTCCCATCTCCGACCTGAATGGATGCCCTTCTGGGTAAATCAAGC 501
121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTrpIysGlnAsnGlyThrIleVal 140
502 GCTGCTCTGCTTTTGGAGGGAATTGATGATGATGATGATGATGATGATGATGATGAT 561
141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleValTrpValIleGlnAsp 160
562 CAAGTAGCAACTATATCAAGAAACATGTTTCAACCAATGGCAAACTGGGTGAACAGAGC 621
161 AsnGlnThrGlnGlyIleTyrGlnThrTrpAsnValIysAlaSerProGlnIleGlyAla 180
622 AATGAAACAGCAATTATATATGATGATGATGATGATGATGATGATGATGATGATGAT 681
181 GlnThrTrpPheAspSerTyrAspCysSerIysPheValIleuArgThrPheAsnIleu 200
682 GAGACATGGTTGATTCCGACGACTGTTCCAAATTTGGTAAAGACCTTTAACAAGTTG 741
201 AlaGlnPheGlyAlaGlnPheIysAsnIleGlnThrAsnGlyTrpArgGilePheIleTyr 220
742 GCTGAATTTGGAGCAGAGTTCAGCAACATAGAAACCACTATACAGAAATTTCTTAC 801
221 SerGlyGlnProThrTyrIleuGlyAsnGlnThrSerValPheGlyProThrGlyAsnIys 240
802 AGTCAGAACCTACTTATCTGGGAATGAACATCTGTTTGGGCCCAACAGAAACAG 861
241 ThrIleuGlyLeuAlaIleIysArgPheTyrTrpProIleIysProHisIleuProThrIys 260
862 ACTCTTGTTTGCCATMAAAAGATTTATTAACCCCTTCAACCACTTTGCCAATMAA 921
261 GlnPheIleuSerIleuGlnIlePheAspAlaValIleValHisIleGlnPheTyr 280
922 GAATTTCTGTGATCTCTTGCATAATTTTATGATGATGATGATGATGATGATGATGAT 981
1041
281 LeuPheTyrAsnPheGlnTyrTrpPheIleuProMetIysPheProPheIleIleThr 300
982 TGTGTTTATATATTTGAATATTTGGTTTATCTGATGAATATTCCTTTATTTAAATACA 1041
301 TyrGlnGlnIleProIleuProIleuArgAsnIysThrIleSerGlyIleu 316
1042 TATGAGAAATCCCTTACCTTACCAAAACAAACCACTCTCGGTTTA 1089

				PRI 03-SEP-2002
LOCUS	AK075109	2120 bp	mRNA	linear
DEFINITION	Homo sapiens cDNA FL390628 f1s, clone PLACE1003407, highly similar to Homo sapiens putative transmembrane protein (CLN5) mRNA.			
VERSION	AK075109			
DESCRIPTION	AK075109.1 GI:22760993			
FEATURES	ORF mapping; f1s (full insert sequence).			
CURSOR	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T., Sugiyama, T., Suzuki, Y., Nagai, K., Sugeno, S., Ishii, S., Kawai, Hio, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Kojima, S., Nagahara, K., Masuhio, Y., Ono, T., Okano, K., Yoshikawa, Y., Aoebuka, S., Saeki, N., Hattori, A., Okumura, K., Iwayanagi, T. and Nihomiya, K.			
TITLE	NEDD human cDNA sequencing project			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 2120)			
AUTHORS	Isogai, T. and Otsuki, T.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomic@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)			
COMMENT	NEDD human cDNA sequencing project supported by Ministry OF			

FEATURES

Economy, Trade and Industry of Japan: cDNA full insert sequencing; Research Association for Biotechnology: cDNA library construction; Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center: cDNA 5' - & 3' end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.).

Location/Qualifiers

```

source
1. .2120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACB1003407"
/tissue_type="placenta"
/clone_id="PLACB1"
/note="cloning vector: pME185FL3

```

fullinsert sequencing:
A library construction:
Tokyo, Laboratory of
5'- & 3'-end one pass
Arch Institute (supported

Alignment Scores:	
Pred. No.:	1,27e-168
Score:	1744.00
Percent Similarity:	99.37%
Best Local Similarity:	99.37%
Query Match:	99.20%
DB:	9
	Gaps: 0

US-10-010-050A-2 COPY 31 346 (1-316) X AK075109 (1-2120)

QY	I	ArgValSerGlyIleProSerArgArgHisTrpProValProTyrLysArgPheAspPhe	20
Db	142	CGGGCTCTGGGGACATCCCTCCCGGGCCACACCGCGGGCCCTCGACACGGCTTTACATTC	201
QY	21	ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro	40
Db	202	CGTCAAAACCTGCATCTTATTGTTCACACTAAGTACTTTCTGTGCCAATGGCTCACCCT	261
QY	41	IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrp	60
Db	262	ATCCACAGTTATGGAGGGGTGATGATGACACTGAAGTTTTCGATTCACAGCCACAGATGG	321
QY	61	GluPheLysTyrGlyAspLeuLeuGlyHisIleuLysIleMetHisAspAlaIleGlyPhe	80
Db	322	GAAATTTAAATAATGGAGACCTCCTGGGACACTTGAAATTAATGCAAGATGCCATTGGATTC	381
QY	81	ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly	100
Db	382	ACAAAGTACATTACTGGCGAAGAACACACAAACAAATGAAATGGATTTCCAACTTGGC	441
QY	101	AsnCystrhrPheProHisIleuArgProGluMetAspAlaProPheTrpCysAsnGlnGly	120
Db	442	AACGTGACATTTCCCATCTCCGACCTCGAAATGGATGCCCCCTTCTGGTGATACAAAGGC	501
QY	121	AlaAlaCysPhePheGlnGlyIleAspArgValHisTrpLysGluAsnGlyThrIleuVal	140
Db	502	GGTGCTGCTTTTGTGAGGAAATGTATGATGTCTACTCGAAGAAATATGGACATTAGTT	561
QY	141	GlnValAlaThrTlIeSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsp	160
Db	562	CAAGTGCACACTATATCAAGAAACATGTTCAACCAAAATGGCAAAAGGGGTGAACAACGAGC	621
QY	161	AsnGluThrGlyIleTyrTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAla	180
Db	622	AATGAACAACGGAATTTATATAGACATGGAAATGTMAAAGCCAGCCAGAAAGGGGGGCA	681
QY	181	GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu	200
Db	682	GACACATGGTTTCACTTCCACAGCTCTTCCAAATTTGTGTAAAGACCTTTAACAGATTG	741
QY	201	AlaGluPheGlyValaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr	220
Db	742	GCTGAATTTGGAGCAAGATTCAAAGAAACAAACCAATCATCAAGAAATATTTCTTTAC	801
QY	221	SerGlyGluProThrTyrIleuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys	240

b	802	AGTGGAGAACCTTACTATTCGGGAAATGAAACATCGTGTGTTGGGCCAACAGAAACAG	86c
y	241	ThirLeuGlyLeuAlaIleLysArgPheLysTrpProPheLysProHisLeuProThrLys	260
b	862	ACTCTTGTTTACCGCATTAACAAAAAGATTTTATTCACCCCTTCAACACCATTTGCCAACTTAAA	921
y	261	GlnPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyr	283
b	922	GAATTCCTGTTGAGTCTCTTGGCAAAATTTTGGATTCAGAGCATTTGTGCAAAACAGTTCTAT	981
y	281	LeuPheTyrAsnPheGluTyrTrpPheLeuProMetLysPheProPheIleLysIleThr	303
b	982	TTCGTTTATTAATTTTGAATATTTGTTTTCACATGAAATTCCTTTATTAATAATACA	1041
y	301	TyrGlnGlnIleProLeuProIleArgAsnLysThrLeuSerGlyLeu	316
b	1042	TATGAAAGAAATCCCTTACCTATCAGAAACAAACACATCTCTGTGTTTA	1089
RESULT 6			
D073413		BD073413	1038 bp
OCUS		Secreted protein which human chromosome 13 encodes.	linear
EPINITION		BD073413	
ACCESSION		BD073413.1	GI:22619016
ERSTION		JP 2001511345-A/12.	
WORDS		JP 2001511345-A/12.	
OURCE		synthetic construct	
ORGANISM		artificial sequences.	
REFERENCE		1 (bases 1 to 1038)	
AUTHORS		Sheppard, P.O. and Gilbertson, D.G.	
TITLE		Secreted protein which human chromosome 13 encodes	
JOURNAL		Patent: JP 2001511345-A 12 14-AUG-2001;	
COMMENT		ZYMOGENETICS INC	
OS		Artificial Sequence	
PN		JP 2001511345-A/12	
PD		14-AUG-2001	
PF		24-JUL-1998	JP 2000504249
PR		24-JUL-1997	US 60/053613
PI		PAUL O SHEPPARD, DIBRA G GILBERTON	
PC		CI2N15/09, A6IK38/00, A6IK48/00, C07K14/47, C07K16/19, CI2N1/15, PC	
		CI2N1/19,	
PC		CI2N1/21, CI2N5/10, CI2P21/02, CI2Q1/68, CI2N15/00, A6IK37/02, CI2N5/	FC
		00	
CC		Degenerate nucleotide sequence encoding zsig46 polypeptide cf	
CC		SRQ ID NO:2	
CC		Key	Location/Qualifiers
FT		source	1..1038
FT		Location/Qualifiers	/organism='Artificial Sequence'.
FEATURES			
source		1..1038	
		/organism="synthetic construct"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:32630"	
ORIGIN			
Alignment Scores:			
Pred. No.:	1,49e-145	Length:	1038
Score:	1515.00	Matches:	267
Percent Similarity:	85.03%	Conservative:	0
Best Local Similarity:	85.03%	Mismatches:	47
Query Match:	86.18%	Indels:	0
DB:	6	Gaps:	0
US-10-010-050A-2_COPY_31_346	(1-316) x BD073413	(1-1038)	
QY	2	ValSerGlyTlleProSerArgArgHisIstrProValProTyrLysArgPheAspPheArg	21
Db	94	GTWMSGGNAATCCCMWSNMGNMGNCAATGGCCGTCNCNTATTAARMGNNTTGAATTMGN	153
QY	22	ProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerProIle	41
Db	154	CNAAACCGAATCANTATATGTCARGCNAARTRATACNTTTCGACNAAGGWSNCCNAATH	213

[illegible]

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heieh, F.,
Ditchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Donald, M.F., Casavant, T.L.,
Schaefer, T.E., Brownstein, M.J., Uedlin, T.B., Tothiyuki, S.,
Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.U.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McEwan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huly, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Keltman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smailus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257
12477932
2 (bases 1 to 2318)
Strausberg, R.
Direct Submission
Submitted (06-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbcr@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nri.nih.gov
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dierich, N.L., Granite, S., Guan, J., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,
Maduro, O.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripp, S., Thomas, P.J., Touchman, J.W.,
Tsungue, C., Vogt, J.D., Walker, M.A., Wecherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H., and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAP Plate: 56 Row: C Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

Location/Qualifiers
1..2318
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5251891"
/tissue_type="Mammary tumor. MAP-TGF alpha model. 7 months
old, gross tissue."
/clone_1lb="NCI CGAP_Mams"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
1..2318
/gene="Clns"
/db_xref="LOCUSID:211286"
/db_xref="MGI:2442253"
1..1013
/gene="Clns"
/codon_start=3
/product="Clns protein"

/protein_id="AAH25487.1"
/db_xref="GI:19343791"
/db_xref="LOCUSID:211286"
/db_xref="MGI:2442253"
/translation="TTPAHRRPALALALGLATLIGASPTNSGQRPVYKRFSPRPKT
DPTQQAITYCPYSPPLPVKMDVLEVLQAPYEFKRGDLGKKHMDAYGRFS
TLGKNTYTIEMVLPOLGNTFPHLRDAPKAPFQWNGAACPFEGIDDKMKENGTLIS
VVATISGNTFNKIAEWKQDNENGIYEVETVRSQDQWTFESYDCSFWFRTYK
KLAFGEFFKLENTYTKIPLYSEPIYLNENSTIFGPKGKTLALIKFVGEPRPY
LSTKDFLMLNLIKFDYTIHROFLFNFWYFPLPMKDPFKITYEETPLPTRTTFT
DL"

ORIGIN

Alignment Scores:
Pred. No.: 1,12e-135 Length: 2318
Score: 1423.00 Matches: 246
Percent Similarity: 89.03% Conservative: 30
Best Local Similarity: 79.35% Mismatches: 34
Query Match: 80.94% Indels: 0
DB: 10 Gaps: 0

US-10-010-050a-2_COPY_31_346 (1-316) x BC025487 (1-2318)

QY 7 SerArgArgHisITRPProValProTyllyARgPheAspPheAqProlyProAspPro 26
DB 81 TCTGGGCAACGCTGGCGGTCCTTACACAGCGCTTCTTCCGTCCGACAGACAGATCC 140
QY 27 TyrcysGlnAlaIstYrThrPheCyeprothrglyserProileProValImetGlnly 46
DB 141 TACGTCAAGTAAAGTAACTTGTGCTTACCGGCTGCGCCACCCAGTTATGAGAC 200
QY 47 AspAspAspIleGluValAPheArgLeuGlnAlaProValITrgIuphelyeTyrclyAsp 66
DB 201 AATGACGATCGACGAGCTTACGACTACACACCCGATTTGGAAATTAATATGAGAC 260
QY 67 LeuleuGlnHisIleuLySileMetHisAspAlaIleGlyPheArgSerThleuThrgly 86
DB 261 CTCCTGGGACACTTAAACTATGATGATGACCGCGTGATTCAGAGACACACTGACAGGC 320
QY 87 LysAsnYrThrMetGluTrrPtyGluIuphlegIuGlnGlyAsnCyAsThPheProHis 106
DB 321 AAGAACTACACAAATGATGATGATGAACTTTCCAGCTGGGCACTGTACATTTCCCCAC 380
QY 107 LeuArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaIaCyAsPheBheGlu 126
DB 381 CTCGGGCTGACAGAGCGCTCCCTTCGTGCTGTAACCAAGGGGCGAGCTGCTTTTGA 440
QY 127 GlyIleAspAspValHisITrPlysgIuAsnGlyThrIleuValGlnValAlaThrIleSer 146
DB 441 GGAATAGATGATTAACACTGGAAGGAAACCGGACACTGTGCACTGTGCAACCATATCC 500
QY 147 GlyAsnMetPheAsnGlnMetAlaIstYrVallysgIuAspAsnGluThrGlyIleTy 166
DB 501 GGAACACATTTAAACAAAGTGCAGAGTGGGTGAGGACAGCAATGAACTGGGATTTAT 560
QY 167 TyrcGluThrTrrPAsnValIleAlaSerProGluIuysGlyAlaGluThrTrrPheAspSer 186
DB 561 TATGAGATATGACGCGCTCCGAGCGCGCCACAGACAAAGGGGCCAACAAGGTTGACATCC 620
QY 187 TyrcAspCySerLysePheValIleuArgThrPheAsnLySleuAlaGluPheGlyAlaGlu 206
DB 621 TACGACTGTTCGAATTTTGTCTTAAGACATATGAAGAAATTTGGCGAATTTGGACAGAA 680
QY 207 PheLyAsnIleGluThrAsnYrThrArgIlePheIuYrSerGlyGluProThrTy 226
DB 681 TTCAAGAGATGAAACAAACATACGAAATAATTTCTTACAGTGGAGAGCTATTTC 740
QY 227 LeuGlyAsnGluThrSerValPheGlyProThrglyAsnLySthLeuGlyLeuAlaIle 246
DB 741 CTGGGAATTAAGAAACATCTATTTTGGGCCAAAGAAACAAAGACTCTTGCCCATTA 800
QY 247 LybArgPheTyrrTyrrProPheLyserProHisIleuPProThrIlysgIuPheIleuSerIleu 266

```

Db      801 AAAAAATTTATGCCCCCTTCAGACCGTATTTGTCAACCAAGATTTCGTGATTC 860
2y      267 LeuGlnIlePheAspAlaValIleValHslyGlnPheTyrLeuPheTyrAsnPheGlu 286
Db      861 TTTGAAATTTTGTGATTCAGTATTAACACAGACAGTCTACTGTTTATTAACCTTGAG 920
2y      287 TTTTTPheLeuPrometLysPheProPheIleLeuTyrTyrGlnGluIleProLeu 305
Db      921 TATTGTTTCTACCATGAAACCCCTTTGTCAAAATACATACAGAAACCCCGTTA 980
2y      307 ProIleArgAsnLysThrLeuSerGlyLeu 316
Db      981 CCTACCCGACATACACATTACCGACTTG 1010

RESULT 8
AC109554/c
LOCUS
DEFINITION
AC109554
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, X., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 198524)

Worley, K. C.

Direct Submission

Submitted (05-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 198524)

Rat Genome Sequencing Consortium.

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 11, 2002 this sequence version replaced gi:21738217. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

Center project name: GQUR

Center clone name: CH230-331D3

----- Summary Statistics -----

Assembly program: Phrap; version 0.990329

Consensus quality: 177481 bases at least Q40

Consensus quality: 179640 bases at least Q30

Consensus quality: 161141 bases at least Q20

Estimated insert size: 183954; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 198524: contig of 198524 bp in length.

Location/Qualifiers

1. 198524

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-331D3"

1. 1183

/note="wgs_end_extension"

clone_end:17"

131403..131443

/note="clone_boundary"

clone_end:17"

site:Mbol

end_sequence:XXAC14TU"

FEATURES

Source

misc_feature

misc_feature

REFERENCE 2 (bases 1 to 227920)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 227920)
 REFERENCE Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 13, 2003 this sequence version replaced gi:22855855.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GIBR
 Center clone name: CH230-13816
 ----- Summary Statistics

Assembly program: Atlas 3.0;
 Consensus quality: 214528 bases at least Q40
 Consensus quality: 21733 bases at least Q30
 Consensus quality: 219064 bases at least Q20
 Estimated insert size: 223690; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length
 (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 223559: contig of 223559 bp in length
 * 223560 223659: gap of unknown length
 * 223660 224738: contig of 1079 bp in length
 * 224739 224838: gap of unknown length
 * 224839 226289: contig of 1451 bp in length
 * 226290 226390: gap of unknown length
 * 226390 227920: contig of 1531 bp in length.

FEATURES
 source 1..227920
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-13806"
 1..1145
 misc_feature
 /note="wgs contig"
 222227..223559
 misc_feature
 /note="wgs_contig"

ORIGIN
 Alignment Scores: 1..7e-124 Length: 227920
 Pred. No.: 1339.00 Matches: 235
 Score:

Percent Similarity: 89.35% Conservative: 25
 Best Local Similarity: 80.76% Mismatches: 31
 Query Match: 76.17% Indels: 0
 DB: 2 Gaps: 0

US-10-010-050a-2_COPY_31_346 (1-316) x AC107510 (1-227920)

QY	7	SeArGAGhIstRPrOvalPrOTrYlYsArGpHeAsPpHeArGPrOTrYpPrOAsPpRO	26
DB	37327	JCTGGGCAAGCGCTGGCCCGTACCTACCAAGCGCTTCCTTCCTCGCAAGCAATGCC	37386
QY	27	TyrGySgInAlaYtYrThrPhcYpPrOThrGlySerPrOIlPrOValMetGluGly	46
DB	37387	TTCTGTCAAGCCAAATGATACCTTCTGCTCTGCTGCTCACTCCACTTGAAGAGC	37446
QY	47	AsPAsPpIlleGlValPheArGLeuGInAlaPrOValTrPGluPheYsTyrgIAsP	66
DB	37447	AATGACGTATGAACTCTTAAGCTACAGCCCGCTTTGGGAATTAAATATGAGAGC	37506
QY	67	LeuLeuGlyAlaLeuYsIlleMetHisAsPAlalleGlyPheArGSerThrLeuThrGly	86
DB	37507	CTCCTGGACACTTCAAAATCATGACGATGCTGATTCAGAGTACGATCGTACCGGAC	37566
QY	87	lysAsnTYrThrMetGluTrpTYrGluLeuPheGlnLeuGlyAsnCyThrPheProHis	106
DB	37567	AAGAATCAACAATTAATGATGATGAGCTTTCCAGCTCGGCACTGTACATTTCCCCAC	37626
QY	107	LeuArGPrGluMeArSpAlaPrOphETrPCYsAsnGlnGlyAlaAlaCySphePheGlu	126
DB	37627	CTTCGCGCGGAAGTGAACGCTCGCTCTGTGTATACAGAGGGGACGCTTCTTTGAA	37686
QY	127	GlyIleAsPAsPValHisTrpYlYsGluAsnGlyThrLeuValGlnValAlaThrIleSer	146
DB	37687	GGATTTACGATTAACACACTGAAAGAAACGGGACCTGTCTGTGTGACCACTATCC	37746
QY	147	GlyAsnMeCPheAsnGlnMetAlaYsTrpValYsGlnAsPasnGluThrGlyIleTy	166
DB	37747	GGAAACACGTTTAAACAAAGTGGCCGAGTGGTGAAGACGACATGAGACTGGGATTAT	37806
QY	167	TyrGluThrTrpAsnValYsAlaSerPrOgluYsGlyAlaGluThrTrpPheAsPser	186
DB	37807	TACGAGCATGACAGCTCGCGGCGACGCCAGAAAGGGCGGACATGTTGAATCC	37866
QY	187	TyrAsPcySerIysPheValIleuArGThrPheAsnYsLeuAlaGluPheGlyAlaGlu	206
DB	37867	TACGACTGTTGCAATTTTGTCTTAAGGACATGAGAAATTTGGCTGATTTGAAACGAA	37926
QY	207	PheIysAsnIleGluThrAsnTYrThrArGIllePheLeuTYrSerGlyGluProThrTy	226
DB	37927	TTCAAGAAAGATGAAACAAACTATATGAAATATTTCTTTACAGCGGAGAAACGATTAC	37986
QY	227	LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnYsThrLeuGlyLeuAlaIle	246
DB	37987	TTGGGAATGAATGAAGTATATTTTGGGCCCAAGAAAGAAACCTTGTGCTTGGCATA	38046
QY	247	LysArGpHeTYrTYrProPheIysProHisLeuProThrIysGluPheLeuSerIeu	266
DB	38047	AAAAAATTTTACGCGCCCGCTCAACCGATTCGTCAACCAAGATTTTCTGTAAATTC	38106
QY	267	LeuGlnIlePheAsPAlaValIleValHisYsGlnPheTYrLeuPheTYrAsnPheGlu	286
DB	38107	TTGAAATTTTTCACACAGTATATATGACAGAGAGTTCTACCGTTTATTAACCTTGA	38166
QY	287	TyTrpPheLeuProMetIysPheProPheIle	297
DB	38167	TATTTGTTTCTACCTATGAACCCCTTGTCTC	38199
RESULT 10			
AC001226	AC001226	106988 bp	DNA linear
LOCUS	AC001226		
DEFINITION	AC001226	Genomic sequence from Human 13, complete sequence.	
ACCESSION	AC001226.1	GI:2133862	
VERSION	AC001226.1	GI:2133862	

WORDS HTG.
 HOMO sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 106988)
 AUTHORS Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W., Fasmann, K.H., and Lander, E.S.
 TITLE Genomic sequence from Human 13
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 106988)
 AUTHORS Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W., Fasmann, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P., Barna, N., Brown, K., Cooke, P., Daly, M.J., Forrest, C., Frisby, W.J., Gage, D., Geraghty, K., Hagos, B., Jacotot, L., Lane, M., Mackenzie, J., Margulis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J., Strickland, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and Zody, M.
 TITLE Direct Submission
 JOURNAL Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 3 (bases 1 to 106988)
 AUTHORS Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W., Fasmann, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P., Barna, N., Chang, A., Cooke, P., Daly, M.J., Forrest, C., Frisby, W.J., Gage, D., Geraghty, K., Hagos, B., Jacotot, L., Lane, M., Mackenzie, J., Margulis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J., Strickland, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and Zody, M.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On May 29, 1997 this sequence version replaced gi:1932720. The Staden databases, finishing information, and all chromatographic files used in the assembly of this clone are available from our anonymous ftp site.

FEATURES
 source All repeats were identified using RepeatMasker: Smt, A.F.A. & Green, P. (1996-1997) Washington.edu/RM/RepeatMasker.html.
 Location/Qualifiers
 1..106988
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone_id="76N15"
 /clone_lib="unknown"
 complement(113..200)
 repeat_region /rpt_family="THR1B"
 complement(202..611)
 repeat_region /rpt_family="L1PA3"
 complement(674..694)
 repeat_region /rpt_family="AT_rich"
 complement(835..981)
 repeat_region /rpt_family="MIR2"
 1632..1920
 repeat_region /rpt_family="AluJo"
 1970..2271
 repeat_region /rpt_family="AluSp"
 2779..2946
 repeat_region /rpt_family="FAM"
 3552..3828
 repeat_region /rpt_family="AluJb"
 4361..4640
 repeat_region /rpt_family="AluSg"
 5253..5385
 repeat_region /rpt_family="MLT1F"
 5914..6106
 repeat_region /rpt_family="MLT1F"
 complement(6341..6435)
 repeat_region

repeat_region /rpt_family="MER34"
 complement(6838..7140)
 repeat_region /rpt_family="AluSx"
 7394..7455
 repeat_region /rpt_family="MER5A"
 7464..7627
 repeat_region /rpt_family="MER5B"
 complement(7624..7792)
 repeat_region /rpt_family="MER5B"
 complement(8363..8391)
 repeat_region /rpt_family="AT_rich"
 8480..8639
 repeat_region /rpt_family="MER5B"
 complement(9199..9286)
 repeat_region /rpt_family="(TA)n"
 9340..9644
 repeat_region /rpt_family="AluJo"
 9880..9946
 repeat_region /rpt_family="MIR2"
 10208..10787
 repeat_region /rpt_family="L1MB3"
 10984..11297
 repeat_region /rpt_family="AluSg"
 11311..11470
 repeat_region /rpt_family="L1PA8"
 complement(11471..11495)
 repeat_region /rpt_family="AT_rich"
 complement(12394..12440)
 repeat_region /rpt_family="AT_rich"
 13782..13921
 repeat_region /rpt_family="MIR"
 14035..14316
 repeat_region /rpt_family="AluSx"
 complement(15018..15046)
 repeat_region /rpt_family="AT_rich"
 15080..15373
 repeat_region /rpt_family="AluJb"
 15568..15881
 repeat_region /rpt_family="AluJo"
 complement(18038..18071)
 repeat_region /rpt_family="(CAT)n"
 complement(20389..20494)
 repeat_region /rpt_family="FAM"
 21238..21534
 repeat_region /rpt_family="AluSx"
 complement(23331..23353)
 repeat_region /rpt_family="GC_rich"
 complement(23546..23578)
 repeat_region /rpt_family="GC_rich"
 24565..24707
 repeat_region /rpt_family="MIR"
 26013..26047
 repeat_region /rpt_family="MIR2"
 complement(27717..28532)
 repeat_region /rpt_family="L1PB1"
 complement(28553..28854)
 repeat_region /rpt_family="AluSg"
 complement(28860..29076)
 repeat_region /rpt_family="L1"
 29365..29561
 repeat_region /rpt_family="MIR"
 complement(30050..30147)
 repeat_region /rpt_family="MIR"
 31487..31812
 repeat_region /rpt_family="AluJo"
 complement(33629..33732)
 repeat_region /rpt_family="MIR"
 34434..34729
 repeat_region /rpt_family="AluSx"
 35070..35361
 repeat_region /rpt_family="AluSx"
 complement(36437..36461)
 repeat_region /rpt_family="AT_rich"

```

repeat_region      complement (39967, .40261)
                    /rpt_family="AluSx"
repeat_region      complement (40741, .40915)
                    /rpt_family="AluSg"
repeat_region      40963, .41663
                    /rpt_family="L1PA2"
repeat_region      42042, .42352
                    /rpt_family="MLTIE"
repeat_region      42467, .42771
                    /rpt_family="AluSg"
repeat_region      42772, .42848
                    /rpt_family="L1MB8"
repeat_region      42845, .43151
                    /rpt_family="Aluub"
repeat_region      43152, .43318
                    /rpt_family="L1MA10"
repeat_region      43253, .43427
                    /rpt_family="L1MB8"
repeat_region      43429, .43672
                    /rpt_family="L1MB8"
repeat_region      complement (44188, .44481)
                    /rpt_family="AluSc"
repeat_region      44544, .45246
                    /rpt_family="LTR8"
repeat_region      complement (46999, .47032)
                    /rpt_family="AT_rich"
repeat_region      complement (47411, .47431)
                    /rpt_family="AT_rich"
repeat_region      complement (47612, .47684)
                    /rpt_family=" (CA) n"
repeat_region      complement (47946, .48239)
                    /rpt_family="Aluub"
repeat_region      complement (48827, .49872)
                    /rpt_family="AT_rich"
repeat_region      complement (50160, .50186)
                    /rpt_family="AT_rich"
repeat_region      50352, .51121
                    /rpt_family="MBE8A"
repeat_region      complement (52140, .52284)
                    /rpt_family="Aluub"
repeat_region      53772, .53974
                    /rpt_family="MER1B"
repeat_region      54225, .54348
                    /rpt_family="MIR"
repeat_region      complement (54533, .54579)
                    /rpt_family=" (TGA) n"
repeat_region      complement (54580, .54736)
                    /rpt_family="Aluub"
repeat_region      complement (54755, .54782)
                    /rpt_family=" (CA) n"

Alignment Scores:
Pred. No.:          1,51e-82          Length:          106988
Score:              925.50             Matches:          173
Percent Similarity: 89.69%             Conservative:    1
Jest Local Similarity: 89.18%           Mismatches:      1
Query Match:        52.65%             Indels:          19
De:                 9                  Gaps:            1

US-10-010-050A-2_COPY_31_346 (1-316) x ACC001226 (1-106988)
QY      123 CysPhepGluGluYlLeaspApyAlHistrPlysgluanglyThrluValGlnVal 142
Db      31851 TGTITTTTAA----- 31862
QY      143 AlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIySTrPVallyGlnAspAsnGlu 162
Db      31863 -----CTAGGAACATGTTCAACCAATGCGAAGTGGTGAAACAGACATGAA 31913
QY      163 ThrGlyIleTyTrGlnThrTrpAsnValysAlaserProGluIyGlyValIagIuThr 182
Db      31914 ACAGAAATTATTATGAGACATGGAATGTAAAGCCAGCCGAAAGGGGGCAGAGACA 31973

```

```

QY      183 TrpPheaspSerTyrraPpCyesserlyspPheValleuArgTrpPheAsnlyLeuAlaGlu 202
Db      31974 TGGTTTGAATTCCTACGACTGTTCCAAATTGTGTGTAAGACCTTTACAGAGTTGGCTGAA 32033
QY      203 PheGlyIaGluPheLyAsnIleGluThrAsnTyrrTharGlyIlePheLeuTySerGly 222
Db      32034 TTGGAGACAGAGCTTCAGAACATAGAACCAACATATACAGAAATATTTCTTTACAGTGA 32093
QY      223 GluProThyTyLeuGlyAsnGluThrservalPheGlyProThrGlyAsnlyThrlu 242
Db      32094 GAACCTACTTATCTGGGAAATGAAACATCTGTTTGGGCCCAAGGAAACAAAGCTCTT 32153
QY      243 GlyLeuAlaIleLyAsnGlyPheTyrrProPheLyProHlsleuProThrlsGluPhe 262
Db      32154 GGTTTACCATTAATAAAGATTATTAATACCTTCACCAACCATTTGCCAATTAAGATTT 32213
QY      263 LeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHlslyGlnPheTyrrleuPhe 282
Db      32214 CTGTTGAGCTCTTCACAAATTTTGTATGACAGATGATGACAAACAGCTTATTTGTTT 32273
QY      283 TyrraPheGluTyrrTyrrPheLeuProMetlyspPheProPheIleLyIleThrTyrrGlu 302
Db      32274 TATTAATTTGAATATGTTGTTTACCTATGAAATTCCTTTATTAATAATTAACATATGAA 32333
QY      303 GluIleProLeuProIleAluAsnlyThrluSerGlyLeu 316
Db      32334 GAAATCCCTTACCTATGAGAAACAAACACTCTCTGTTTA 32375

RESULT 11
AL136440      169362 bp      DNA      linear      HTG 10-JUL-2001
LOCUS      Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.
DEFINITION      AL136440
ACCESSION      AL136440.2 GI:10039473
VERSION      HTG; HTGS PHASE1; HTGS CANCELLED.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
TITLE      Direct Submission
JOURNAL      Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
              requests: clonerequest@sanger.ac.uk
              On Sep 8, 2000 this sequence version replaced gi:6982057.
              ----- Genome Center
              Center: Sanger Centre
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquerry@sanger.ac.uk
              ----- Project Information
              Center project name: B165124
              ----- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Sequencing vector: plasmid; 108752; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Consensus quality: 162085 bases at least Q40
              Consensus quality: 164521 bases at least Q30
              Consensus quality: 166040 bases at least Q20
              Insert size: 168062; sum-of-contigs
              Insert size: 164996; 1.7% error; agarose-fp
              Quality coverage: 4.08x in Q20 bases; sum-of-contigs Quality
              coverage: 4.16x in Q20 bases; agarose-fp

***** NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

[illegible]

Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridge, UK. E-mail enquiries: humquerry@sanger.ac.uk
CB30 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonemaster@sanger.ac.uk

On Aug 22, 2000 this sequence version replaced gi.8894433.

Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk

Project Information

Center project name: bA453N2

Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 220597 bases at least Q40

Consensus quality: 234299 bases at least Q30

Consensus quality: 241103 bases at least Q20

Insert size: 245967; sum-of-contigs

Quality coverage: 2.52x in Q20 bases; sum-of-contigs Quality

coverage: 3.92x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      2130: contig of 2130 bp in length
*      2131      2230: gap of 100 bp
*      2231      5689: contig of 3459 bp in length
*      5689      5789: gap of 100 bp
*      5789      10045: contig of 4257 bp in length
*      10045      10146: gap of 100 bp
*      10146      12624: contig of 2478 bp in length
*      12624      12724: gap of 100 bp
*      12724      20993: contig of 8269 bp in length
*      20993      21093: gap of 100 bp
*      21093      28935: contig of 7843 bp in length
*      28935      29037: gap of 100 bp
*      29037      32230: contig of 3194 bp in length
*      32230      32331: gap of 100 bp
*      32331      35153: contig of 2823 bp in length
*      35153      35154: gap of 100 bp
*      35154      37649: contig of 2396 bp in length
*      37649      37650: gap of 100 bp
*      37650      40245: contig of 2496 bp in length
*      40245      40345: gap of 100 bp
*      40345      48420: contig of 8075 bp in length
*      48420      48521: gap of 100 bp
*      48521      52169: contig of 3649 bp in length
*      52169      52269: gap of 100 bp
*      52269      56443: contig of 4174 bp in length
*      56443      56543: gap of 100 bp
*      56543      63088: contig of 6545 bp in length
*      63088      63089: gap of 100 bp
*      63089      65281: contig of 2093 bp in length
*      65281      65381: gap of 100 bp
*      65381      67940: contig of 2559 bp in length
*      67940      68040: gap of 100 bp
*      68040      75029: contig of 6989 bp in length
*      75029      75130: gap of 100 bp
*      75130      85567: contig of 10438 bp in length
*      85567      85668: gap of 100 bp
*      85668      89150: contig of 3483 bp in length
*      89150      89250: gap of 100 bp
*      89250      93380: contig of 4130 bp in length
*      93380      93480: gap of 100 bp
*      93480      95619: contig of 2139 bp in length
*      95619      95719: gap of 100 bp
*      95719      100582: contig of 4873 bp in length
*      100582      100693: gap of 100 bp
*      100693      103959: contig of 3267 bp in length

```

FEATURES

source

```

*      103960      104059: gap of 100 bp
*      104060      110871: contig of 6812 bp in length
*      110872      110971: gap of 100 bp
*      110972      115555: contig of 4584 bp in length
*      115556      115655: gap of 100 bp
*      115656      118593: contig of 2938 bp in length
*      118594      118693: gap of 100 bp
*      118694      123530: contig of 4837 bp in length
*      123531      123630: gap of 100 bp
*      123631      127087: contig of 3457 bp in length
*      127088      127187: gap of 100 bp
*      127188      129684: contig of 2497 bp in length
*      129685      129784: gap of 100 bp
*      129785      132132: contig of 2348 bp in length
*      132133      132332: gap of 100 bp
*      132333      143792: contig of 11560 bp in length
*      143793      143892: gap of 100 bp
*      143893      148766: contig of 4874 bp in length
*      148767      148866: gap of 100 bp
*      148867      152358: contig of 3392 bp in length
*      152359      152358: gap of 100 bp
*      152359      156922: contig of 4564 bp in length
*      156923      157022: gap of 100 bp
*      157023      160153: contig of 3131 bp in length
*      160154      160253: gap of 100 bp
*      160254      164334: contig of 4081 bp in length
*      164335      164335: gap of 100 bp
*      164335      166000: contig of 2166 bp in length
*      166001      166700: gap of 100 bp
*      166701      171896: contig of 5196 bp in length
*      171897      171996: gap of 100 bp
*      171997      184915: contig of 12919 bp in length
*      184916      185015: gap of 100 bp
*      185016      187238: contig of 2223 bp in length
*      187239      187338: gap of 100 bp
*      187339      189602: contig of 2264 bp in length
*      189603      189702: gap of 100 bp
*      189703      194563: contig of 4761 bp in length
*      194564      194563: gap of 100 bp
*      194564      197682: contig of 3119 bp in length
*      197683      197782: gap of 100 bp
*      197783      201788: contig of 4006 bp in length
*      201789      201888: gap of 100 bp
*      201889      204060: contig of 2172 bp in length
*      204061      204160: gap of 100 bp
*      204161      206237: contig of 2077 bp in length
*      206238      206337: gap of 100 bp
*      206338      210376: contig of 4039 bp in length
*      210377      210476: gap of 100 bp
*      210477      225245: contig of 14769 bp in length
*      225246      225345: gap of 100 bp
*      225346      230027: contig of 4682 bp in length
*      230028      230127: gap of 100 bp
*      230128      232610: contig of 2483 bp in length
*      232611      232710: gap of 100 bp
*      232711      234942: contig of 2232 bp in length
*      234943      235042: gap of 100 bp
*      235043      240958: contig of 5916 bp in length
*      240959      241058: gap of 100 bp
*      241059      251187: contig of 10129 bp in length.

```

Location/Qualifiers

1. 251187

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="13"

/clone="RP11-453N22"

/clone_id="RP11-453N22"

1. 2130

/note="assembly_fragment:00469"

fragment_chain:1"

2231. 5689

/note="assembly_fragment:01219"

misc_feature

misc_feature

misc_feature

[illegible]

Qy	243	GlyLeuAlaIleIysArgPheTyrTyrProPheIysProHisIleuProThylsGluPhe	262
Db	193804	GGTTACGCATTAATAAAGATTATTATTACCCCTTCAACACCACTTTGGCCAACTTAAGAAATTT	193745
Qy	263	LeuIleuSerIleuLeuGlnIlePheAspAlaValIleValHisIysGlnPheTyrIleuPhe	282
Db	193744	CTGTTGGAGTCTCTGGCAAAATTTTGAATGACATGATTTGGACAAACAGTTTCATTTGTTT	193685
Qy	283	TyrAsnGheIuTyrTyrPheLeuProMetIysPheProPheIleIysIleTyrGlu	302
Db	193684	TATATTTTGAATATTGTTTTCACATATGAAATTCCTTTTATTAATAATACATATGAA	193625
Qy	303	GluIleProIeuProIleArgAsnIysThrIeuSerGlyIeu	316
Db	193624	GAATTCCTTACCTATCAGAAACAAACACTCTCTGTTTA	193583
RESULT 13			
BD125187		BD125187	697 bp DNA linear PAT 18-SEP-2002
LOCUS			
DEFINITION			Primer for synthesizing full-length cDNA and use thereof.
ACCESSION		BD125187	
VERSION		BD125187.1	GI:23220132
KEYWORDS		JP 2002017375-A/618.	
SOURCE			
ORGANISM			Homo sapiens (human)
REFERENCE			
AUTHORS			Homo sapiens
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
			Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
			1 (bases 1 to 697)
			Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
			Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
			Koga,H.
			Primer for synthesizing full-length cDNA and use thereof
TITLE			Patent: JP 2002017375-A 618 22-JAN-2002;
JOURNAL			HELIX RESEARCH INSTITUTE
COMMENT			
			OS Homo sapiens (human)
			PN JP 2002017375-A/618
			PD 22-JAN-2002
			PF 07-JUL-2000 JP 2000253172
			PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
			PI ISHII,
			PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
			SHINICHI KOJIMA,
			PI TETSUJI OTSUKI,HISASHI KOGA
			PC
			C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
			10',C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
			Primer for synthesizing full-length cDNA and use thereof FH Key
FEATURES			
source			
			FT source
			FT
			Location/Qualifiers
			1..697
			/organism='Homo sapiens (human)'
			/organism='Homo sapiens'
			/mol_type='genomic DNA'
			/db_xref='taxon:9606'
ORIGIN			
			Alignment Scores:
			Pred. No.:
			Score:
			Percent Similarity:
			Best local Similarity:
			Query Match:
			DB:
			US-10-010-050A-2_COPY_31_346 (1-316) x BD125187 (1-697)
Qy	1	ArgValSerGlyIleProSerArgArgHisTyrProValProTyrIysArgPheAspPhe	20
Db	142	CGGGTCTCGGAGCATCCCTCCCGCGGCATATGCGCGGTGCGCTTCGACGCGCTTGACTTC	201

ORIGIN

Alignment Scores:

Pred. No.:	2,386-82	Length:	697
Score:	986.00	Matches:	170
Percent Similarity:	92.51%	Conservative:	3
Best Local Similarity:	90.91%	Mismatches:	8
Query Match:	50.97%	Indels:	6
DB:	6	Gaps:	2

US-10-010-050A-2_COPY_31_346 (1-316) x BD126453 (1-697)

QY	1	ArgValSerGlyIleProSerArgArgHisTrpProValProTyrLysArgPheAspPhe	20
Db	142	CGGGTCCTCGGCACTCCCTCCCGGCGCACTGGCCGGTCCCTCGAGCGCTTACCTC	201
QY	21	ArgProLysProAspProTyrCySGlnAlaLysTyrThrPheCySProThGlySerPro	40
Db	202	CGTCCAAACCTGATCTCTTATTCGACGCTAAGATACCTTCTGTCGCAACTGGCTCACC	261
QY	41	IleProValMetGluGlyAspAspAspTleGluValPheArgLeuGlnAlaProValTTrp	60
Db	262	ATCCACAGTTATGGAGGGGATGATGACACCTGAAGTTTTCGATTACAGCCAGCTATGG	321
QY	61	GlnPheLysTyrGlyAspPheLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe	80
Db	322	GAATTAAATGAGAGACCTCTCGGACACATTGAAATTTATGATGATGACATTCGGATTG	381
QY	81	ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrrPyrGluLeuPheGlnLeuGly	100
Db	382	AGAGTACATTACATCGGACAGACACACCAATGCAATGGTATGAACTTTCCAACTTGGC	441
QY	101	AsnCyThrPheProHisLeuLysArgProGluMetAspAlaProPheTrrCyAsnGlnGly	120
Db	442	AACGTACATTTCCCATCTCCGACCTGAAAGATGCCCTTCTGGTGATATCAAGC	501
QY	121	AlaAlaCySPhePheGlnGlyTleAspAspValHisTrrPheGlnLysGlnLysThrLeu	140
Db	502	GCTGCTGCTTTTGGAGGAATTGATGATCTTCACTGGAAGGAAATGGGACATTAACT	561
QY	140	IglnValAlaThrIleSer-GlyAsnMetPheAsnGlnMetAla-LysTrrPValLysGln	155
Db	562	TCAATGAGCACTATATATCAAGGAAACATGTTCAACCAATGGCAAAAGTGGGAAACAA	621
QY	160	--AspAsnGluThrGlyLysTyrGlu-ThrTrpAsnValLysAlaSerProGluTrr	175
Db	622	GGACATATGAAACAAGAAATTATATATGAGACATGAAATTTTAAAGCCANCCANMA	681
QY	178	S---GlyAlaGln 181	
Db	682	AAANGGGGCAAAA 694	

RESULT 15
AE014175.0
WPCOMMENT
Sequence split into 4 fragments LOCUS AE014175 Accession AE014175
Fragment Name Begin End
AE014175_0 1 110000
AE014175_1 100001 210000
AE014175_2 200001 310000
AE014175_3 300001 404829
AE014175.3 404829 bp DNA linear ROD 06-AUG-2001
DEFINITION
Mus musculus p16hald deletion region section 3 of 11 of the complete sequence.
ACCESSION
AE014175 AE013600
VERSION
AE014175.1 GI:22128029
KEYWORDS
SOURCE
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 404829)
 Peterson, K., King, B., Hagg-Greenberg, A., Roik, J., Bull, C. and
 O'Brien, T.
 Functional and comparative genomic analysis of the piebald deletion
 region of mouse chromosome 14
 Genomics 80 (2), 172 (2002)
 JOURNAL 22150871
 MEDLINE 12160731
 PUBMED 2 (bases 1 to 404829)
 AUTHORS Adams, M. and Mural, R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAY-2002) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 The strings of n's in this record represent gaps between contigs,
 and the length of each string corresponds to the length of the gap
 except where there are 20 n's. A string of 20 n's represents a gap
 of unknown length.

FEATURES
 source
 1..404829
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="14"
 /map="Mit164-Mit165"
 <38137..>47216
 /gene="Irf1"
 <38137..>47216
 /gene="Irf1"
 /product="Immune responsive gene 1"
 /note="partial transcript sequence derived from alignment
 to Genbank Accession Number L38281"
 <61259..>68607
 /gene="Cln5"
 join(<61259..61421,62721..62886,64274..64499,
 66916..>68607)
 /gene="Cln5"
 /product="ceroid-lipofuscinosis, neuronal 5"
 /note="partial transcript derived from alignment to
 tentative consensus sequence TC227789, from TIGR Mouse
 Gene Index (v. 6.0)."
 complement(<73858..>86476)
 /gene="Fbl3a"
 complement(join(<73858..74501,80293..80464,83295..83417,
 86126..>86476))
 /gene="Fbl3a"
 /product="leucine-rich repeat-containing F-box protein
 Fbl3a"
 complement(join(73858..74501,80293..80464,83295..83417,
 86126..>86476))
 /codon_start=1
 /product="leucine-rich repeat-containing F-box protein
 Fbl3a"
 /protein_id="AA092567.1"
 /db_xref="GI:22128030"
 /translation="RMKGRDSDSDSAEKGTAEKPRPTTQERSOPCDMGNLQDI
 VHVFRYLPLIDRAHSQVCRNMQVFHMDLRCFEPINOPATSYLAKTHPELTKO
 IIRKSNHLYVSPKVSKEAACAIDLSQVNCSLKTLGLISTARSPMDLPKSH
 FIALTFVFNRSKSLKIDDPVDDPSLKLVANNSDTLKLRSSCPHVSFAGIL
 CVADOCHELEALNYHLSDLLALISSEKHYLEHLRIDVVENSGQTHFTICKS
 SNAFLIKHSPKXNLVNFYFLEPDPFRYEIPATHLYGFSKDVLAGRVGTCPS
 LVELVVCANGRLPDLBELIRIARCKNLISIGICEVSCSAFVERFYKMGGRSLGCLS
 IMEEVLIPDQKYSLEQIHMEVSKHGRWPFDMPTW"
 complement(<106136..>339446)
 /gene="Pbr1"
 complement(<106136..>339446)
 /gene="Pbr1"
 /product="Pam/hw/rpm-1"
 /note="PAM, protein associated with Myc; partial
 transcript derived from alignment of Homo sapiens NCBI
 RefSeq sequence NM_015057"

ORIGIN

Alignment Scores:
 Pred. No.: 1.7e-62 Length: 110000
 Score: 730.00 Matches: 127
 Percent Similarity: 87.65% Conservative: 22
 Best Local Similarity: 74.71% Mismatches: 21
 Query Match: 41.52% Indels: 0
 DB: 10 Gaps: 0

US-10-010-050a-2_COPY_31_346 (1-316) x AE014175_0 (1-110000)

QY	147	GLYAsmehcPheasnGlnMetAlaValTyrValIysGlnAspAengIunrGlyIleTyr	166
Db	66915	GGAAACACATTTCACAAAGTGGCCGAGTGGTGAAGCAGACATGAACCTGGATTAT	66974
QY	167	TyrGIunrTyrPasnValIysAlaSerProGluIysGlyAlaGIunrTyrPheaspSer	186
Db	66975	TATAGACATGAGACGGTCCGAGCCGCCAGACAGACAGGGGCCAGAGCTGTGTGAGTCC	67034
QY	187	TyrAspCyserIysPheValIleuArgThrPheasnIysIleuAlaGIuPheGlyAlaGIu	206
Db	67035	TACGACTGTTCGAATTTGTCTTAAGACATVTAAGAAATGGCTGAAATTTGAGACAGAA	67094
QY	207	PheIysAsnIleGIunrAsnTyrThrArgIlePheIeuTyrSerGIuPProThrTyr	226
Db	67095	TTCAAGAGATGAACAACTATACGAATATTTCTTACATGAGAGACCTATTAC	67154
QY	227	IeuGIyAsnGIunrSerValPheGIyProThrGIyAsnIysThrIeuGIyIleuAlaIle	246
Db	67155	CTGGGAATGAACATCATTTTGGGCCCAAGAGAAACAAAGACTCTTGGCCATA	67214
QY	247	LysArgPheIyTyrProPheIysProHisIeuProThrIysGIuPheIeuSerIeu	266
Db	67215	AAAAAATTTTATGGCCCTTCAAGCCGATTTGTCAACCAAGATTTCTGATGAATTC	67274
QY	267	IeuGIuIlePheAspAlaValIleValHisIysGIuPheIyIeuPheTyrAsnIleGIu	286
Db	67275	TTGAAATTTTGTATACGTGATTTATACAGACAGCTCTACTGTTTATATACCTTGAG	67334
QY	287	TyrTyrPheIeuProIleIysPheProPheIleIyIleThrTyrGIuGIuIleProIeu	306
Db	67335	TATGTGTTTCAACCAATGAACCCCTTGTGCAAAATACATACGAAGAAACCCGTTTA	67394
QY	307	ProIleArgAsnIysThrIeuSerGIyIeu	316
Db	67395	CCTACCCGACATACAAACATTTACGACCTTG	67424

Search completed: April 25, 2004, 03:23:00
 Job time : 6261.28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame_plus_p2n model

1 on: April 24, 2004, 22:59:32 ; Search time 529.278 Seconds
(without alignments)
2292.945 Million cell updates/sec

file: US-10-010-050A-2_COPY_1_28
effect score: 154
sequence: 1 MRRGAGAGRAGSACWALMLMLAVYPG 28

oring table: BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

arched: 3470272 segs, 2167151695 residues
tal number of hits satisfying chosen parameters: 640544

num DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

MODEL=frame+p2n.model -DEV=xlh
=/cgn2_1/USPTO_SPOOL/US10010050/runat_22042004_113204_27549/app_query_fastq_1.1372
B=genemb1 -QMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
NITS=bits -START=1 -END=-1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45
LOCALIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
JUTPUT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US10010050 @CGN 1.1 7433 @runat_22042004_113204_27549 -NCPU=6 -ICPU=3
NO MAP -LARGEST=NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1: *
1: gb_ba: *
2: gb_hg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_scs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_in: *
18: em_hum: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_scs: *
28: em_un: *

29: em_vl: *
30: em_hg_hum: *
31: em_hg_inv: *
32: em_hg_other: *
33: em_hg_mus: *
34: em_hg_pln: *
35: em_hg_rod: *
36: em_hg_mem: *
37: em_hg_vit: *
38: em_sy: *
39: em_hgo_hum: *
40: em_hgo_mus: *
41: em_hgo_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	154	100.0	BD125187	BD125187 Primer fo
2	154	100.0	BD126453	BD126453 Primer fo
3	154	100.0	BD073402	BD073402 Secreted
4	154	100.0	BD136338	BD136338 95 human
5	154	100.0	BD127858	BD127858 Primer fo
6	154	100.0	AK075109	AK075109 Homo sapi
7	154	100.0	AF068227	AF068227 Homo sapi
8	154	100.0	AC001226	AC001226 Genomic s
9	154	100.0	AL136440	AL136440 Homo sapi
10	146	94.8	AL358875	AL358875 Homo sapi
11	106	68.8	BD073413	BD073413 Secreted
12	73	47.4	MI067213	MI067213 Mesothelio
13	72	46.8	HSBA19D2	AL080248 Human DNA
14	71	46.1	AC079116	AC079116 Homo sapi
15	71	46.1	AC008177	AC008177 Homo sapi
16	71	46.1	AC007248	AC007248 Homo sapi
17	70.5	45.8	AC087615	AC087615 Homo sapi
18	70.5	45.8	AC074349	AC074349 Homo sapi
19	70.5	45.8	AF075622	AF075622 Sequence
20	69	44.8	AF070499	AF070499 Daphnia m
21	69	44.8	AK120197	AK120197 Oryza sat
22	69	44.8	AR385625	AR385625 Sus scrofa
23	69	44.8	AC117471	AC117471 Homo sapi
24	69	44.8	AL137007	AL137007 Human DNA
25	69	44.8	AP003906	AP003906 Oryza sat
26	69	44.8	AP004589	AP004589 Oryza sat
27	69	44.8	AC079984	AC079984 Homo sapi
28	69	44.8	AL161716	AL161716 Human DNA
29	69	44.8	AC110302	AC110302 Mus muscu
30	69	44.8	AP005442	AP005442 Oryza sat
31	69	44.8	SC0939115	AL939115 Streptomy
32	68	44.2	AC108168	AC108168 Homo sapi
33	68	44.2	AL606530	AL606530 Human DNA
34	68	44.2	AC091025	AC091025 Homo sapi
35	67.5	43.8	PM282B	AL684841 Penicillini
36	67.5	43.8	AL596209	AL596209 Mouse DNA
37	67.5	43.8	AC113407	AC113407 Homo sapi
38	67.5	43.8	AC021965	AC021965 Homo sapi
39	67.5	43.8	AC084044	AC084044 Mus muscu
40	67.5	43.8	CHPMPCA	L29028 Chlamydomon
41	67	43.5	AL627434	AL627434 Human DNA
42	67	43.5	AC025259	AC025259 Homo sapi
43	67	43.5	AC116479	AC116479 Mus muscu
44	67	43.5	AL442068	AL442068 Homo sapi
45	66.5	43.2	HS328178	AJ328178 Homo sapi

RESULT 1

ALIGNMENTS

LOCUS	BD125187	697 bp	DNA	linear	PAT 18-SEP-2002
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BD125187				
VERSION	BD125187.1	GI:23220132			
KEYWORDS	JP 2002017375-A/618.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Oca,T., Nishikawa,T., Isegai,T., Hayashi,K., Ishii,S., Kawai,Y., Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.				
TITLE	Primer for synthesizing full-length cDNA and use thereof				
JOURNAL	Patent: JP 2002017375-A 618 22-JAN-2002;				
COMMENT	HELIX RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002017375-A/618 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISEGAI,KOJI HAYASHI,SHIZUKO PI YURI YAMAI,AI MAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUJI OTSUKI,HISASHI KOGA PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC 10',C12P21/02,C12Q1/68/C12B21/08,G06F17/30,C12M15/00,C12M5/00 CC Primer for synthesizing full-length cDNA and use thereof FH key Location/Qualifiers FT source 1..697 FI location/Qualifiers 1..697 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"				
FEATURES					
source					
ORIGIN					
Alignment Scores:					
Pred. No.:	1..14e-08	Length:	697		
Score:	154.00	Matches:	28		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatch:	0		
Query Match:	100.00%	Indels:	0		
DB:	6	Gaps:	0		
US-10-010-050A-2_COPY_1_28 (1-28) x BD125187 (1-697)					
Qy	1 MetatgaTgGtAlaAgTAlaAlaTgGtAlaTgcTTCgtCTTpaLaLeuAlaLeu 20				
Ddb	52 ATGGCGGCGGGGCGCGGCGCGGCGGTTCGTGAGGAGCGCGCTTCGTGAGGCGCTGCGCTG 111				
Qy	21 LeuTTPLeuAlaValValProgly 28				
Ddb	112 CTTGAGCTGGCGGCGGTTCGGGC 135				
RESULT 2					
LOCUS	BD126453	697 bp	DNA	linear	PAT 18-SEP-2002
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BD126453				
VERSION	BD126453.1	GI:23221398			
KEYWORDS	JP 2002017375-A/1884.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Oca,T., Nishikawa,T., Isegai,T., Hayashi,K., Ishii,S., Kawai,Y., Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and				

[illegible]

CC Secreted protein which human chromosome 13 encodes FH Key
 FT CDS Location/Qualifiers
 (47) . (1084) .
 source 1.1486
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

IGIN

Alignment Scores:
 ad. No.: 2.22e-08 Length: 1486
 re: 154.00 Matches: 28
 Percent Similarity: 100.00% Conservative: 0
 at Local Similarity: 100.00% Mismatches: 0
 Indels: 0
 Gaps: 0

10-010-050A-2_COPY_1_28 (1-28) x BD073402 (1-1486)

1 MetAArgGlyAlaGlyAlaAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeu 20
 47 ATGCGGGGGGGCGGGCGGGCGGCTCGGGGACGGCTTCCTGAGCTGGCGCGCTG 106
 21 LeuTrpLeuAlaValAlaProGly 28
 107 CTTGGCTCGGGGTGGTCCGGGC 110

SUIT 4
 136338 BD136338 1751 bp DNA linear PAT 18-SEP-2002
 CUS 95 human secretory proteins.
 DEFINITION BD136338
 BD136338.1 GI:23231283
 REVISION UP 2002506627-A/25.
 YWORDS Homo sapiens (human)
 URCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1751)
 Ruben, S.M., Ni, Y., Rosen, C.A., Yu, G.L., Young, P.E., Fen, P.,
 Soppet, D.R., Wei, Y.F., Endress, G.A., Duan, R.D., Kyaw, H., Ehner, R.,
 Lafleur, D.W., Olsen, H.S., Shi, Y. and Moore, P.A.
 95 human secretory proteins
 Patent: JP 2002506627-A 25 05-MAR-2002;
 JOURNAL HUMAN GENOME SCIENCES INC
 COMMENT OS Homo sapiens (human)
 PN UP 2002506627-A/25
 PD 05-MAR-2002
 PR 18-MAR-1999 UP 2000536733
 PR 19-MAR-1998 US 60/078566, 19-MAR-1998 US 60/078576 PR
 19-MAR-1998 US 60/078573, 19-MAR-1998 US 60/078574 PR
 19-MAR-1998 US 60/078579, 19-MAR-1998 US 60/078578 PR
 19-MAR-1998 US 60/078581, 19-MAR-1998 US 60/078577 PR
 19-MAR-1998 US 60/078583, 01-APR-1998 US 60/080314 PR
 01-APR-1998 US 60/080312, 01-APR-1998 US 60/080313 PI STEVEN
 M RUBEN, JIAN NI, CRAIG A ROSEN, GDO
 LIANG YU PAUL, E YOUNG,
 PI PING FENG,
 PI DANIEL R SOPPET, YING FEI WEI, GREGORY A ENDRESS, ROXANNE D DUAN,
 PI HUA KYAW,
 PI REINHARD EHNER, DAVID W LAFLUR, HENRIK S OLSEN, YANGSU SHI, PAUL,
 PI A MOORE
 PC C12N15/09, A61K38/00, A61K48/00, A61P43/00, C07K14/47, C07K16/18,
 PC C12N1/15,
 PC C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12O1/68, G01N3/53, G01N3/56,
 PC C12N15/00, A61K37/02, C12N5/00
 CC n equals a,t,g, or c
 CC n equals a,t,g, or c
 CC n equals a,t,g, or c
 CC n equals a,t,g, or c

CC n equals a,t,g, or c
 CC n equals a,t,g, or c
 FH Key Location/Qualifiers
 FT source 1.1751
 /organism="Homo sapiens (human)"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

FEATURES
 source 1.1751
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
 Pred. No.: 2.56e-08 Length: 1751
 Score: 154.00 Matches: 28
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0

US-10-010-050A-2_COPY_1_28 (1-28) x BD136338 (1-1751)

QY 1 MetAArgGlyAlaGlyAlaAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeu 20
 49 ATGCGGGGGGGCGGGCGGGCGGCTCGGGGACGGCTTCCTGAGCTGGCGCGCTG 108
 DB 21 LeuTrpLeuAlaValAlaProGly 28
 DB 109 CTTGGCTCGGGGTGGTCCGGGC 132

RESULT 5
 BD127858 2120 bp DNA linear PAT 18-SEP-2002
 LOCUS BD127858
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD127858
 BD127858.1 GI:23222803
 VERSION UP 2002017375-A/3289.
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2120)
 Oca, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 3289 22-JAN-2002;
 JOURNAL HELIX RESEARCH INSTITUTE
 COMMENT OS Homo sapiens (human)
 PN UP 2002017375-A/3289
 PD 22-JAN-2002
 PF 07-JUL-2000 UP 2000253172
 PI TOSHIO OCA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI ISHII,
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUO OCA, HISASHI KOGA
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,
 PC C12P21/02, C12O1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 FT CDS Location/Qualifiers
 (16) . (1089) .
 source 1.2120
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
 Pred. No.: 3.03e-08 Length: 2120

Note: 154.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

1-10-010-050A-2_COPY_1_28 (1-28) x BD127858 (1-2120)

1 MetAargtgglyalaglyalalalargtgglyalasertrpocystrpaleuallaleu 20
52 ATGCGGCGGGGCGCGGCGGCGGCTGCGGGAACGGCTTCTGCTGCTGCGGCGCTG 111
21 LeuTripleuAlaValValProGly 28
112 CTTTGGCTCGCGGCGGTTCGGGCG 135

RESULT 6 2120 bp mRNA linear PRI 03-SEP-2002
CDS AK075109 Homo sapiens CDNA FLJ90628 fls. clone PLACE1003407, highly similar
to Homo sapiens putative transmembrane protein (CIN5) mRNA.

ACCESSION AK075109
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
Kawai-Hio,Y., Saito,K., Yamamoto,D., Makamatsu,A., Nakamura,Y.,
Kojima,S., Nagahara,K., Masuo,Y., Ono,T., Okano,K., Yoshikawa,Y.,
Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
Ninomiya,K.

REFERENCE NEDO human cDNA sequencing project
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2120)
TITLE Isogai,T. and Otsuki,T.

REFERENCE Direct Submission
JOURNAL Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
Institute of Medical Science, University of Tokyo, Laboratory of
Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
sequencing and clone selection: Helix Research Institute (supported
by Japan Key Technology Center etc.).

FEATURES Location/Qualifiers
SOURCE 1..2120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1003407"
/tissue_type="placenta"
/clone_id="PLACE1"
/note="Cloning vector: pME18SFL3"

ORIGIN

Alignment Scores: 3.03e-08 Length: 2120
Pred. No.: 154.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

JS-10-010-050A-2_COPY_1_28 (1-28) x AK075109 (1-2120)

1 MetAargtgglyalaglyalalalargtgglyalasertrpocystrpaleuallaleu 20

Db 52 ATGCGGCGGGGCGCGGCGGCGGCTGCGGGAACGGCTTCTGCTGCTGCGGCGCTG 111
Qy 21 LeuTripleuAlaValValProGly 28
112 CTTTGGCTCGCGGCGGTTCGGGCG 135

RESULT 7 4080 bp mRNA linear PRI 28-JUL-1998
LOCUS AF068227 Homo sapiens putative transmembrane protein (CIN5) mRNA, complete
cde.

ACCESSION AF068227
VERSION AF068227.1 GI:3342385
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 4080)
Lander,E.S., Sawukoski,M., Klockars,T., Holmberg,V., Santavuori,P., Lander,E.S.
and Peltonen,L.
CIN5, a novel gene encoding a putative transmembrane protein
mutated in Finnish variant late infantile neuronal ceroid
lipofuscinosis
Nat. Genet. 19 (3), 286-288 (1998)

JOURNAL Nat. Genet. 19 (3), 286-288 (1998)
MEDLINE 96324783
PUBMED 9662406
REFERENCE 2 (bases 1 to 4080)
AUTHORS Sawukoski,M., Klockars,T., Holmberg,V., Santavuori,P., Lander,E.S.
and Peltonen,L.

REFERENCE Direct Submission
JOURNAL Submitted (26-MAY-1998) Department of Human Molecular Genetics,
National Public Health Institute, Mannerheimintie 166, Helsinki
00300, Finland

FEATURES Location/Qualifiers
SOURCE 1..4080
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q22"
1..4080
/gene="CIN5"
1293..2516
/gene="CIN5"
/codon_start=1
/product="putative transmembrane protein"
/protein_id="AAC27614.1"
/db_xref="GI:3342386"

ORIGIN
Alignment Scores: 5.39e-08 Length: 4080
Pred. No.: 154.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-010-050A-2_COPY_1_28 (1-28) x AF068227 (1-4080)

Qy 1 MetAargtgglyalaglyalalalargtgglyalasertrpocystrpaleuallaleu 20
Db 1476 ATGCGGCGGGGCGCGGCGGCGGCTGCGGGAACGGCTTCTGCTGCTGCGGCGCTG 1535

21 LeuTRpleu1aValValProgly 28
 |||||
 1536 CTTTGCTCGCGGTGTTCCGGGC 1559

SUITE 8
 CUS 001226 AC001226 106988 bp DNA linear PRI 29-MAY-1997
 FINITION Genomic sequence from Human 13, complete sequence.
 CESSION AC001226
 REGION AC001226.1 GI:2133862
 TWORDS HTG.
 TPRODS
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 (bases 1 to 106988)
 HAWKINS, T.L., REEVE, M.P., CHRISTOFFERSEN, A., BIRREN, B.W.,
 PASMAN, K.H., LANDER, E.S., MCKERNAN, K., MUNRO, C., RICHARDSON, P.,
 BARN, N., BROWN, K., COOKE, P., DALY, M.J., FORREST, C., FRIPP, W.J.,
 GAGE, D., GARAIGERY, K., HAGOS, B., JACOTOT, L., LANE, M., MACKENZIE, J.,
 MARGIS, N., MCDERMOTT, J., MOLONEY, N., MORROW, J., NACHMAN, A.,
 NAYLOR, J., O'CONNOR, T., PETERSON, K., ROLLINS, G., SPENCER, J.,
 STILLWELL, J., STONE, C., STRICKLAND, C., SYDNEY, K., WILMER, F. and
 ZODY, M.
 Direct Submission
 Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 106988)
 HAWKINS, T.L., REEVE, M.P., CHRISTOFFERSEN, A., BIRREN, B.W.,
 PASMAN, K.H., LANDER, E.S., MCKERNAN, K., MUNRO, C., RICHARDSON, P.,
 BARN, N., BROWN, K., COOKE, P., DALY, M.J., FORREST, C., FRIPP, W.J.,
 GAGE, D., GARAIGERY, K., HAGOS, B., JACOTOT, L., LANE, M., MACKENZIE, J.,
 MARGIS, N., MCDERMOTT, J., MOLONEY, N., MORROW, J., NACHMAN, A.,
 NAYLOR, J., O'CONNOR, T., PETERSON, K., ROLLINS, G., SPENCER, J.,
 STILLWELL, J., STONE, C., STRICKLAND, C., SYDNEY, K., WILMER, F. and
 ZODY, M.
 Direct Submission
 Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 29, 1997 this sequence version replaced gi:1932720.
 The Staden databases, finishing information, and all
 chromatographic files used in the assembly of this clone are
 available from our anonymous ftp site.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 Location/Qualifiers
 1. 106988
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="76N15"
 /clone_id="unknown"
 /complement(113..200)
 /rpt_family="THERB"
 /complement(202..611)
 /rpt_family="L1PA3"
 /complement(674..694)
 /rpt_family="AT_rich"
 /complement(835..981)
 /rpt_family="MIR2"
 1632..1920
 /rpt_family="AluB"
 1970..2271
 /rpt_family="AluBp"
 2779..2946
 repeat_region

/rpt_family="FAM"
 3552..3928
 /rpt_family="AluB"
 4361..4640
 /rpt_family="AluSg"
 5253..5385
 /rpt_family="MLT1F"
 5914..6106
 /rpt_family="MLT1F"
 /complement(6341..6435)
 /rpt_family="MER34"
 /complement(6838..7140)
 /rpt_family="AluSx"
 7394..7455
 /rpt_family="MER5A"
 7464..7627
 /rpt_family="MER5B"
 /complement(7624..7792)
 /rpt_family="MER5B"
 /complement(8363..8391)
 /rpt_family="AT_rich"
 8480..8639
 /rpt_family="MER5B"
 /complement(9199..9286)
 /rpt_family="(TA)n"
 9340..9644
 /rpt_family="AluB"
 9880..9946
 /rpt_family="MIR2"
 10208..10787
 /rpt_family="L1MB3"
 10984..11297
 /rpt_family="AluSg"
 11311..11470
 /rpt_family="L1PA8"
 /complement(11471..11495)
 /rpt_family="AT_rich"
 /complement(12384..12440)
 /rpt_family="AT_rich"
 13782..13921
 /rpt_family="MIR"
 14035..14316
 /rpt_family="AluSx"
 /complement(15018..15046)
 /rpt_family="AT_rich"
 15080..15373
 /rpt_family="AluB"
 15586..15881
 /rpt_family="AluB"
 /complement(18038..18071)
 /rpt_family="(CA)n"
 /complement(20389..20494)
 /rpt_family="FAM"
 21236..21534
 /rpt_family="AluSx"
 /complement(23331..23353)
 /rpt_family="GC_rich"
 /complement(23546..23578)
 /rpt_family="GC_rich"
 24565..24707
 /rpt_family="MIR"
 26013..26047
 /rpt_family="MIR2"
 /complement(27717..28532)
 /rpt_family="L1PB1"
 /complement(28553..28854)
 /rpt_family="AluSg"
 /complement(28860..29076)
 /rpt_family="L1"
 /rpt_family="MIR"
 /complement(29365..29561)
 /rpt_family="MIR"
 /complement(30050..30147)
 /rpt_family="MIR"
 repeat_region

```

Db      23544  ANTGCGCGG36GCGCGGCGCGGCTCGGGAGCGCGGCTTCTGTGTGCTGGGCGCTTGAGCGCTG  23603
Ox      21    LeuTripleuAlaValIvaIProsly  28
          |||||
          |||||
Db      23604  CTTTGCTCGCGGTGTTCGGGCG  23627

RESULT 9
AL136440
LOCUS    Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.
DEFINITION
ACCESSION
VERSION  AL136440.2  GI:10039473
KEYWORDS  HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiindaes; Homo.

REFERENCE
1        Burton,T.
          Direct Submission
          Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
          requests: clonerequests@sanger.ac.uk
          On Sep 8, 2000 this sequence version replaced gi:6982057.

COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

----- Project Information
Center project name: BA185124

----- Summary Statistics
Assembly program: XGAP; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 162085 bases at least Q40
Consensus quality: 166040 bases at least Q20
Insert size: 168062; sum-of-contigs
Insert size: 164996; 1.7% error; agarose-fp
Quality coverage: 4.08x in Q20 bases; sum-of-contigs Quality
Coverage: 4.16x in Q20 bases; agarose-fp

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
*      1      3400: contig of 3400 bp in length
*      3401      3500: gap of 100 bp
*      3501      16444: contig of 12944 bp in length
*      16445      16544: gap of 100 bp
*      16545      19514: contig of 2970 bp in length
*      19515      19614: gap of 100 bp
*      19615      22241: contig of 2627 bp in length
*      22242      22341: gap of 100 bp
*      22341      22526: contig of 3185 bp in length
*      22527      25626: gap of 100 bp
*      25627      33070: contig of 7444 bp in length
*      33071      33170: gap of 100 bp
*      33171      67620: contig of 34450 bp in length
*      67621      67720: gap of 100 bp
*      67721      93413: contig of 25663 bp in length
*      93414      93513: gap of 100 bp
*      93514      114081: contig of 20568 bp in length
*      114082      114181: gap of 100 bp
*      114182      126206: contig of 12025 bp in length
*      126207      126306: gap of 100 bp
*      126307      144583: contig of 18277 bp in length
*      144584      144683: gap of 100 bp

```

```

*      552270      56443: contig of 4174 bp in length
*      56444      56543: gap of 100 bp

```


JOURNAL Submitted (02-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 REQUEST: clonerequest@sanger.ac.uk

COMMENT

On Aug 10, 1999 this sequence version replaced g1:5596948. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human Chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-19D2 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-19D2. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-19D2 is at 1 in this sequence. The true right end of clone RP5-988G17 is at 32774 in this sequence.

FEATURES

source

1. 166913

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="20"

/clone="RP11-19D2"

/clone_11b="RP11-11.1"

8. 277

/note="match: STS: Em:G48897"

match: GSS: Em:B81947"

305. 307

/note="clone RP5-988G17"

aca in this entry

substitution"

/replace="ata"

578. 580

/note="clone RP5-988G17"

tat in this entry

substitution"

/replace="tgt"

703. 705

/note="clone RP5-988G17"

gic in this entry

substitution"

/replace="gcc"

932. 934

/note="clone RP5-988G17"

tat in this entry

substitution"

/replace="ctt"

1146. 1530

/note="MSTR repeat: matches 1. 426 of consensus"

1287. 1289

/note="clone RP5-988G17"

gic in this entry

substitution"

/replace="gat"

variation

1331. 1334

/note="clone RP5-988G17"

gic in this entry

insertion"

/replace="gc"

1603. 1791

/note="MER63A repeat: matches 23. 210 of consensus"

1887. 2298

/note="MTR1H repeat: matches 15. 492 of consensus"

2186. 2188

/note="clone RP5-988G17"

cgg in this entry

substitution"

/replace="cag"

2647. 2649

/note="clone RP5-988G17"

at in this entry

substitution"

/replace="act"

3512. 4065

/note="L1M63 repeat: matches 5573. 6150 of consensus"

4493. 4599

/note="MIR repeat: matches 20. 126 of consensus"

5546. 5637

/note="L1M4A repeat: matches 6204. 6295 of consensus"

5893. 5949

/note="MTR1J repeat: matches 118. 174 of consensus"

6259. 6619

/note="match: GSS: Em:AQ748546"

6620. 7171

/note="L2 repeat: matches 1888. 2498 of consensus"

7172. 7225

/note="L2 repeat: matches 1888. 2498 of consensus"

7274. 7371

/note="L2 repeat: matches 1719. 1818 of consensus"

join(8707. 8791, 21988. 22167)

/note="match: GSS: Em:AL015091"

9501. 9602

/note="MTR1J repeat: matches 103. 203 of consensus"

9868. 10251

/note="MTR1A1 repeat: matches 1. 363 of consensus"

10318. 10436

/note="L2 repeat: matches 2163. 2282 of consensus"

10443. 11001

/note="L1P repeat: matches 3907. 4455 of consensus"

11003. 11100

/note="ALUSg/x repeat: matches 198. 293 of consensus"

11113. 14194

/note="L1P3 repeat: matches 479. 3386 of consensus"

14184. 14296

/note="L1P5 repeat: matches 5331. 5439 of consensus"

14328. 15499

/note="L1M4 repeat: matches 3508. 4696 of consensus"

15668. 16085

/note="L1MEC repeat: matches 1913. 2360 of consensus"

16083. 16300

/note="L1ME1 repeat: matches 5894. 6087 of consensus"

17213. 17513

/note="ALUSg repeat: matches 1. 307 of consensus"

17999. 18457

/note="L1PA8 repeat: matches 5705. 6163 of consensus"

18458. 18651

/note="L1PA8 repeat: matches 5509. 5704 of consensus"

19091. 19130

/note="L20 copies 2 mer at 77% conserved"

complement(19216. 19477)

/note="match: GSS: Em:AQ005419"

20104. 20550

/note="match: GSS: Em:AQ246957"

20304. 20306

/note="clone RP5-988G17"

gic in this entry

substitution"

substitution"


```

repeat_region /rpt_family="MIR"
9105. .9309
/rpt_family="L2"
repeat_region /rpt_family="L2"
9315. .9447
/rpt_family="L2"
repeat_region /rpt_family="MIR"
9848. .9894
/rpt_family="MIR"
10033. .10522
/rpt_family="L1"
repeat_region /rpt_family="L1"
10523. .10935
/rpt_family="ERVU"
repeat_region /rpt_family="ERVU"
11051. .11107
/rpt_family="ERVU"
repeat_region /rpt_family="L1"
11109. .14499
/rpt_family="L1"
repeat_region /rpt_family="ERVU"
14500. .14593
/rpt_family="ERVU"
repeat_region /rpt_family="ERVU"
14594. .14945
/rpt_family="ERVU"
repeat_region /rpt_family="L1"
14946. .16084
/rpt_family="L1"
repeat_region /rpt_family="MIR"
16086. .16442
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
16443. .16815
/rpt_family="MIR"
repeat_region /rpt_family="L1"
16816. .17148
/rpt_family="MIR"
repeat_region /rpt_family="L1"
17351. .17355
/rpt_family="L1"
repeat_region /rpt_family="L1"
17498. .17933
/rpt_family="L1"
repeat_region /rpt_family="MIR"
17993. .18179
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
20280. .20344
/rpt_family="MIR"
repeat_region /rpt_family="ERVU"
20349. .20724
/rpt_family="ERVU"
repeat_region /rpt_family="ERVU"
20723. .20765
/rpt_family="Alu"
repeat_region /rpt_family="ERVU"
20786. .20890
/rpt_family="ERVU"
repeat_region /rpt_family="MIR"
20920. .21040
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
21846. .21922
/rpt_family="L2"
repeat_region /rpt_family="MIR"
22488. .22592
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
25690. .25818
/rpt_family="L2"
repeat_region /rpt_family="L2"
26126. .26425
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
26689. .26911
/rpt_family="L2"
misc_feature /note="similar to EST AI316470 (NID:G4031737) u560f10.y1"
30862. .31090
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
31559. .31806
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
31862. .32102
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
33621. .33835
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
34010. .34116
/rpt_family="MIR"
repeat_region /rpt_family="MIR2_type"
34221. .34300
/rpt_family="MIR2_type"
repeat_region /rpt_family="MIR2_type"
35380. .35498
/rpt_family="L2"
repeat_region /rpt_family="L2"
35527. .35572
/rpt_family="L2"
repeat_region /rpt_family="MIR"
38009. .38147
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
38409. .38767
/rpt_family="MIR"

```

```

repeat_region 38946. .39246
/rpt_family="Alu"
repeat_region 39528. .39830
/rpt_family="Alu"
repeat_region 39982. .40376
/rpt_family="L2"
repeat_region 40384. .40514
/rpt_family="Alu"
repeat_region 41336. .41487
/rpt_family="MIR"
repeat_region 42204. .42336
/rpt_family="MIR"
repeat_region 42527. .42744
/rpt_family="MIR"

```

Alignment Scores:

```

Pred. No.: 3.13e+03 Length: 152996
Score: 71.00 Matches: 13
Percent Similarity: 60.87% Conservative: 1
Best Local Similarity: 56.52% Mismatches: 9
Query Match: 46.10% Indels: 0
DB: 9 Gaps: 0

```

US-10-010-050a-2_copy_1_28 (1-28) x AC008177 (1-152996)

```

QY 4 GYALAGLYALAAARGGLYARGALASERTTPCYSTTPALALEUAALEULEUTPLEU 23
DB 105444 GGTCAATGCTGATGACAGAGGTGGCTTCACATGCTGTGGGACGCTCTACTCTATGGCTT 105385
QY 24 AAlaValVal 26
DB 105384 TGCAGGCTA 105376

```

Search completed: April 25, 2004, 03:24:31
Job time : 620.278 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - nucleic search, using frame_plus_p2n model

on: April 24, 2004, 22:59:32 ; Search time 567.083 Seconds

(without alignments)
2292.945 Million cell updates/sec

File: us-10-010-050a-2_copy_1_30

Infect score: 169

Sequence: 1 MRRGAGARGARASWCWALALIMLAVVPGWS 30

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Indexed: 3470272 segs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Set-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODBL=frimex+ p2n.model -DEV=xlh
3=/cgcn2.1/USPTO.spool/US10010050/runat.22042004.113204.27549/app.query.fasta.1.1372
DB=GenEmbl -QPWT=firstp -SUPFIX=rge -MTNMATCH=0.1 -IOOPL=0 -IOOEXT=0
NITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human4.cdi -LIST=45
DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
JUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
JSER=US10010050 @CGN 1.1 7433 @runat.22042004.113204.27549 -NCPU=6 -ICPU=3
NO MAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=5
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_ov:*
- 22: em_or:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	169	100.0	697	6	BD125187	BD125187 Primer fo
2	169	100.0	697	6	BD126453	BD126453 Primer fo
3	169	100.0	1486	6	BD073402	BD073402 Secretd
4	169	100.0	1751	6	BD136338	BD136338 95 human
5	169	100.0	2120	6	BD127858	BD127858 Primer fo
6	169	100.0	2120	9	AK075109	AK075109 Homo sapi
7	169	100.0	4080	9	AF068227	AF068227 Homo sapi
8	169	100.0	106988	2	AL136440	AL136440 Homo sapi
9	169	100.0	169362	2	AL135987	AL135987 Homo sapi
10	161	95.3	251187	2	BD073413	BD073413 Secretd
11	117	63.2	1038	6	PM2H2B	PM2H2B
12	78.5	45.9	28835	9	HS444G9	Z98258 Human DNA s
13	77.5	45.9	256073	9	AE006464	AE006464 Homo sapi
14	77.5	45.9	105779	9	AL137007	AL137007 Human DNA
15	76	44.4	73807	9	AL606530	AL606530 Human DNA
16	75	44.4	52314	9	AL627434	AL627434 Human DNA
17	74	43.8	290625	2	AL442068	AL442068 Homo sapi
18	74	43.8	118955	2	AC068274	AC068274 Homo sapi
19	73	43.2	152050	1	ML0672113	ML0672113 Mesothiz
20	73	43.2	191072	2	AC115775	AC115775 Mus muscu
21	73	43.2	229553	2	AC115773	AC115773 Mus muscu
22	73	43.2	186359	10	AL596209	AL596209 Mouse DNA
23	72.5	42.9	188202	2	AC084044	AC084044 Mus muscu
24	72.5	42.9	100072	9	AL137144	AL137144 Human DNA
25	72	42.6	166913	9	HSBA19D2	AL1080248 Human DNA
26	72	42.3	249773	2	AC103064	AC103064 Rattus no
27	71.5	42.3	321034	2	AC133998	AC133998 Rattus no
28	71.5	42.0	13897	1	ABCLISA	LA1732 Glucanaceco
29	71	42.0	56050	9	AC097719	AC097719 Homo sapi
30	71	42.0	145969	9	AC025028	AC025028 Homo sapi
31	71	42.0	147856	2	AC079116	AC079116 Homo sapi
32	71	42.0	152996	2	AC008177	AC008177 Homo sapi
33	71	42.0	153740	2	AC027196	AC027196 Homo sapi
34	71	42.0	168467	2	AL450346	AL450346 Human DNA
35	71	42.0	174124	2	AC021386	AC021386 Homo sapi
36	71	42.0	184406	2	AC009476	AC009476 Homo sapi
37	71	42.0	201535	9	AC007248	AC007248 Homo sapi
38	71	42.0	344805	1	BX640434	BX640434 Bordetell
39	71	42.0	348074	1	BX640449	BX640449 Bordetell
40	71	42.0	348134	1	BX640420	BX640420 Bordetell
41	71	41.7	759	8	AK102002	AK102002 Oryza sat
42	70.5	41.7	1028	8	AK059403	AK059403 Oryza sat
43	70.5	41.7	56975	2	AC087615	AC087615 Homo sapi
44	70.5	41.7	153732	2	AC132883	AC132883 Mus muscu
45	70.5	41.7	153732	2	AC132883	AC132883 Mus muscu

RESULT 1

ALIGNMENTS

BD125187 697 bp DNA linear PAT 18-SEP-2002

CCUS Primer for synthesizing full-length cDNA and use thereof.

DEFINITION BD125187

ERSTON BD125187

YWORDS JP 2002017375-A/618.

JURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 697)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.

Patient: JP 2002017375-A 618 22-JAN-2002;

HEALTH RESEARCH INSTITUTE

OS Homo sapiens (human)

PN JP 2002017375-A/618

PD 22-JAN-2002

PF 07-JUN-2000 JP 20020253172

PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YORI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA

PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12NE/ PC
10' C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key

Location/Qualifiers

FT source 1..697

FT location/Qualifiers

1..697 /organism='Homo sapiens (human)'.
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

Alignment Scores:

Pred. No.: 2 696-10 Length: 697

Score: 169.00 Matches: 30

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: Gaps: 0

US-10-010-050A_2_COPY_1_30 (1-30) x BD125187 (1-697)

OY 1 MetaArgArgGlyAlaGlyAlaAlaArgGlyAlaSerTTPCySTRPalalaValaLeu 2)

Dd 52 ATGGGGCGGGCGGGCGGGCTGGGGAACGCCGTTCGTGGTGCTGGGACTGGCGCTG 111

OY 21 LeuTRPLeuAlaValaValaProGlyTYRper 30

Dd 112 CTTCGGCTGCAGGCTGCTCCGAGCTGATCC 141

RESULT 2

LOCUS BD126453 697 bp DNA linear PAT 18-SEP-2002

DEFINITION BD126453

ACCESION BD126453

VERSION BD126453.1 GI:23221398

KEYWORDS JP 2002017375-A/1884.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 697)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and

REFERENCE

AUTHORS

TITLE	Koga,H.
JOURNAL	Primer for synthesizing full-length cDNA and use thereof Patent: JP 2002017375-A 1884 22-JAN-2002; HELIX RESEARCH INSTITUTE
COMMENT	OS Homo sapiens (human) PN JP 2002017375-A/1884 PD 22-JAN-2002 PF 07-JUL-2000 JP 20002531172 PI TOSHIO OYA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO I, PI ISHII, PI YORI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUJI OTSUKI,HISASHI KOGA PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC 10, PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key Location/Qualifiers FT source 1..697 /organism='Homo sapiens (human)'. Location/Qualifiers FI 1..697 /organism='Homo sapiens' /mol_type='genomic DNA' /db_xref='taxon:9606'
FEATURES	
source	
ORIGIN	
Alignment Scores:	
Pred. No.:	2.69e-10 Length: 697
Score:	169.00 Matches: 30
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	Gaps: 0
US-10-010-050A-2_COPY_1_30 (1-30) x BD126453 (1-697)	
Oy	1 MetATGATGTGTAAGTGAATAAGTGTAAGAATSerTPPCysTTPTatLeuAlaLeu 20
Dd	52 ATGGCGGGGGGGGGGGGGGGGGCTTCCTTCTGTGCTGGCCCTTGGGGCTG 111
Oy	21 LeuTPLeuAlaValAlProGlyTrpSer 30
Dd	112 CTITGGCTCGGGGGGGTTCGGGGCTGGTCC 141
RESULT 3	
BD073402	1486 bp DNA linear PAT 27-AUG-2002
LOCUS	
DEFINITION	Secreted protein which human chromosome 13 encodes.
ACCESSION	BD073402
VERSION	BD073402.1 GI:22619005
KEYWORDS	JP 2001511345-A/1. Homo sapiens (human)
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 1486) Sheppard,P.O. and Gilberton,D.G. Secreted protein which human chromosome 13 encodes Patent: JP 2001511345-A 1 14-AUG-2001;
AUTHORS	ZYMOGENETICS INC
TITLE	OS Homo sapiens (human)
JOURNAL	PN JP 2001511345-A/1 PD 14-AUG-2001 PF 24-JUL-1998 JP 2000504249 PR 24-JUL-1997 US 60/053613 PI PAUL,O SHEPPARD,DIBRA G GILBERTON PC C12N15/09,A61K38/00,A61K48/00,C07K14/47,C07K16/18,C12N1/15, PC C12N1/19,
COMMENT	PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,C12N15/00,A61K37/02,C12N5/ PC 00

	CC	n equals a,t,g, or c CC n equals a,t,g, or c FH Key location/Qualifiers FT source 1..1751 location/Qualifiers /organism='Homo sapiens (human)' /mol_type='genomic DNA' /db_xref='taxon:9606'
ORIGIN		
Alignment Scores:		
Pred. No.:	6.29e-10 169.00 Percent Similarity: 100.00% Best local Similarity: 100.00% Query Match: 100.00%	Length: 1751 Matches: 30 Conservative: 0 Mismatches: 0 Indels: 0 Gaps: 0
US-10-010-050A-2_COPY_1_30 (1-30) x BD136338 (1-1751)		
OY	1 MetArgArgGIyAlaGIyAlaAlaArgGIyArGIyAserTrpCyStrPAlaleuAlaIeu 20	
Dd	49 ATCGGGGGGGGGGGGGGGGGCTTCTGTGTCGGGCCCTGGACCTG 108	
OY	21 LeuTrIPleuaIAVaIVaIProGLyTPSer 30	
Dd	109 CTTTGCTCGCGGTGTTCGGGGCTGTACC 138	
RESULT 5 BD127858		
LOCUS	BD127858	2120 bp DNA linear PAT 18-SEP-2002
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.	
ACCESSION	BD127858	
VERSION	BD127858.1 GI:23222803	
KEYWORDS	JP 2002017375-A/3289.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2120) Oca,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.	
AUTHORS	Koga,H.	
TITLE	Primer for synthesizing full-length cDNA and use thereof	
JOURNAL	Patent: JP 2002017375-A 3289 22-JAN-2002;	
COMMENT	HEALTH RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002017375-A/3289 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO PI ISHII, PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUJI OTSUKI,HISASHI KOGA	
FEATURES	source	
CDS	Location/Qualifiers 1..2120 /organism='Homo sapiens' /mol_type='genomic DNA' /db_xref='taxon:9606'	
ORIGIN		
Alignment Scores:		
Pred. No.:	7.5e-10	Length: 2120

```

Score:          169.00      Matches:       30
Percent Similarity:    100.00%   Conservative:    0
Best Local Similarity:    100.00%   Mismatches:     0
Query Match:           100.00%   Indels:         0
                                   Gaps:           0
B:                             5

S-10-010-050A-2_COPY_1_30 (1-30) x BD127858 (1-2120)

Y      1 MetcagaagcglyalaglialaalaaargglyaagaiaSerTrrCystrpalaaleualaleu 20
b      52 ATGCGCGGGGGGCGCGGGCGCGGCTCGGGGAACGGGCTTCGTGtgcTgGcccttcggcgctts 111
Y      21 LeutrpleuaAlaValAlpProglYTrrSer 30
b      112 CTTTGcCTCGCGGtGTrTCGGGcCTGcTGC 141

RESULT 6
AK075109             2120 bp      mRNA      linear      PRI 03-SEP-2002
OCUS                Homo sapiens cDNA FLJ90628 fis, clone PLACE1003407, highly similar
DEFINITION          To Homo sapiens putative transmembrane protein (CLNS) mRNA.
ACCESSION            AK075109
VERSION              AK075109.1 GI:22760983
KEYWORDS             oligo capping; fis (full insert sequence).
SOURCE               Homo sapiens (human)
ORGANISM             Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS             Isogai,T., Ota,T., Nishikawa,T., Hayaashi,K., Otaki,T.,
                    Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
                    Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
                    Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y.,
                    Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
                    Nimomiya,K.
TITLE               MEDO human cDNA sequencing project
JOURNAL              Unpublished
REFERENCE            2 (bases 1 to 2120)
AUTHORS              Isogai,T. and Otsuki,T.
TITLE               Direct Submission
JOURNAL              Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
                    Genomics Laboratory, 1532-3 Yana-Kisarazu, Chiba 292-0812, Japan
                    (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT              MEDO human cDNA sequencing project supported by Ministry of
                    Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                    Research Association for Biotechnology; cDNA library construction:
                    Institute of Medical Science, University of Tokyo; Laboratory of
                    Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
                    sequencing and clone selection: Helix Research Institute (supported
                    by Japan Key Technology Center etc.).
                    Location/Qualifiers
FEATURES             source          1..2120
                                /organism="Homo sapiens"
                                /mol_type="mRNA"
                                /db_xref="taxon:9606"
                                /clone="PLACE1003407"
                                /issue_type="Placenta"
                                /clone_lib="PLACE1"
                                /note="Cloning vector: pME18SFL3"

ORIGIN
Alignment Scores:
Pred. No.:          7.5e-10      Length:      2120
Score:              169.00      Matches:      30
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:        100.00%   Indels:         0
DB:                  Gaps:           0

US-10-010-050A-2_COPY_1_30 (1-30) x AK075109 (1-2120)
Y      1 MetcagaagcglyalaglialaalaaargglyaagaiaSerTrrCystrpalaaleualaleu 20

```

Db	52	ATGCGGGCGGGGCGCGGGCCGCCGCTGAGGGACGCCTTCTGTGCTGGAGCCTTCGGCGCTG	111
Qy	21	LeutridleualaValaValProglYtPser	30
Db	112	CTTTGGCTGGCGGTGGTTCCGGGGCTGTCTC	141
RESULT 7			
LOCUS	AF068227	4080 bp	mRNA linear PRI 28-JUL-1998
DEFINITION	Homo sapiens putative transmembrane protein (CLN5)	mRNA, complete cds.	
ACCESSION	AF068227		
VERSION	AF068227.1	GI:3342385	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 4080) Santavuori,M., Klofars,T., Holmberg,V., Santavuori,P., Lander,E.S. and Peltonen,L.		
TITLE	CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis		
JOURNAL	Nat. Genet. 19 (3), 286-288 (1998)		
MEDLINE	98324783		
PUBMED	9662406		
REFERENCE	2 (bases 1 to 4080) Santvuori,M., Kiockars,T., Holmberg,V., Santavuori,P., Lander,E.S. and Peltonen,L.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-MAY-1998) Department of Human Molecular Genetics, National Public Health Institute, Mannerheimintie 166, Helsinki 00300, Finland		
FEATURES			
source	Location/Qualifiers		
	1..4080		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="13"		
	/map="13q22"		
	1..4080		
	/gene="CLN5"		
	1293..2516		
	/gene="CLN5"		
	/codon_start=1		
	/product="putative transmembrane protein"		
	/protein_id="AAC27614.1"		
	/db_xref="GI:3342385"		
	/translation="MRNRIRGPSSSGADAGCAPRPIAAPHMLPPASOASRGSSG TGCSIMAEVDTAQGAERKRGARBARASPCNALALLMTAVDGMRSVGIGPSRRHM VLPKPRPFDRPDPCQAQTFPCPTSPIPVEGDDEIVFRIQAFVWFFPKKGIDLG HKIKMHDAIIGFRSTLTGKNYTWMEYELFOLGNCTPHLRBMDAPFCNGAACTFGG ITDDVMKNGTLVOVATISGMENQMKAWKYQDNDETGIYYETLVNVAKSPKGAETFG SYDCKPYLIRFNLAERGAERKIENNYRIFLYSGEPYILGNETSVPFTQNKTIIG IAIKRFYPFPPLHPTEKERFILSLQITDAIVIHQFLFNFNEFWLFPMKEPFIKITY EELPIRLINKTLSGL"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,37e-09	Length:	4080
Score:	169.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0
US-10-010-050A-2 COPY_1_30 (1-30) x AF068227 (1-4080)			
Qy	1	MetargarglYAlaGlYAlaAlaArggLYArGaLaSeRTPrCySTPrAlaLeuaLaLeu	20
Db	1476	ATGCGGCGGGGCGCGGGCGCGGCTCGGGGACCGCGCTTCTGTGTGTGGAGCCTTCGGCGCTG	1535

21 LeuTripleAlaValProGlyTyrSer 30
 |||||
 1536 CTTGGCTCGCGGTGTTCGCGGCTGCTCC 1565

SUIT 8
 001226

AC001226 106988 bp DNA linear PRI 29-MAY-1997
 AC001226 Genomic sequence from Human 13, complete sequence.
 AC001226
 AC001226.1 GI:2133862

HTG.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS
 Hawkins, K.H., Reeve, M.P., Christoffersen, A., Birren, B.W.,
 Fauman, K.H. and Lander, E.S.
 1 (bases 1 to 106988)
 Genomic sequence from Human 13
 Unpublished

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 2 (bases 1 to 106988)
 Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,
 Fauman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,
 Batra, N., Brown, K., Cooke, P., Daly, M.J., Forrest, C., Frapp, W.J.,
 Gage, D., Geradigery, K., Haggos, B., Jacotot, L., Lane, M., Mackenzie, J.,
 Margulis, N., McDermott, J., Moloney, N., Morrow, U., Nachman, A.,
 Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J.,
 Stillwell, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and
 Zody, M.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Direct Submission
 Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 106988)
 Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,
 Fauman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,
 Batra, N., Chang, A., Cooke, P., Daly, M.J., Forrest, C., Frapp, W.J.,
 Gage, D., Geradigery, K., Haggos, B., Jacotot, L., Lane, M., Mackenzie, J.,
 Margulis, N., McDermott, J., Moloney, N., Morrow, U., Nachman, A.,
 Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J.,
 Stillwell, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and
 Zody, M.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Direct Submission
 Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 29, 1997 this sequence version replaced gi:1932720.
 The Staden databases, finishing information, and all
 chromatographic files used in the assembly of this clone are
 available from our anonymous ftp site.

FEATURES
 source
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 location/Qualifiers
 1..106988
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="76N15"
 /clone_lib="unknown"
 /complement(113..200)
 /rpt_family="THE1B"
 /rpt_family="L1PA3"
 /rpt_family="L1PA3"
 /complement(674..694)
 /rpt_family="AT_rich"
 /complement(835..981)
 /rpt_family="MIR2"
 /rpt_family="L1A20"
 /rpt_family="AluJo"
 /rpt_family="AluJo"
 /rpt_family="AluSp"
 /rpt_family="MIR"
 /rpt_family="MIR"
 /rpt_family="MIR2"
 /rpt_family="L1PB1"
 /complement(2853..28854)
 /rpt_family="AluSg"
 /complement(28860..29076)
 /rpt_family="L1A"
 /complement(29365..29561)
 /rpt_family="MIR"
 /complement(30050..30147)
 /rpt_family="MIR"

repeat_region /rpt_family="FAM"
 3552..3828
 /rpt_family="AluDb"
 repeat_region 4361..4640
 /rpt_family="AluSg"
 repeat_region 5253..5385
 /rpt_family="MLT1F"
 repeat_region 5914..6106
 /rpt_family="MLT1F"
 repeat_region complement(6341..6435)
 /rpt_family="MER34"
 repeat_region complement(6838..7140)
 /rpt_family="AluSx"
 repeat_region 7394..7455
 /rpt_family="MER5A"
 repeat_region 7464..7627
 /rpt_family="MER5B"
 repeat_region complement(7624..7792)
 /rpt_family="MER5B"
 repeat_region complement(8363..8391)
 /rpt_family="AT_rich"
 repeat_region 8480..8639
 /rpt_family="MER5B"
 repeat_region complement(9199..9286)
 /rpt_family="(TA)n"
 repeat_region 9340..9644
 /rpt_family="AluJo"
 repeat_region 9880..9946
 /rpt_family="MIR2"
 repeat_region 10208..10787
 /rpt_family="L1MB3"
 repeat_region 10984..11297
 /rpt_family="AluSg"
 repeat_region 11311..11470
 /rpt_family="L1PA8"
 repeat_region complement(11471..11495)
 /rpt_family="AT_rich"
 repeat_region complement(12394..12440)
 /rpt_family="AT_rich"
 repeat_region 13782..13921
 /rpt_family="MIR"
 repeat_region 14035..14316
 /rpt_family="AluSx"
 repeat_region complement(15018..15046)
 /rpt_family="AT_rich"
 repeat_region 15080..15373
 /rpt_family="AluDb"
 repeat_region 15588..15881
 /rpt_family="AluJo"
 repeat_region complement(18038..18071)
 /rpt_family="(CAT)n"
 repeat_region complement(20389..20494)
 /rpt_family="FAM"
 repeat_region 21238..21534
 /rpt_family="AluSx"
 repeat_region complement(23331..23353)
 /rpt_family="GC_rich"
 repeat_region complement(23546..23578)
 /rpt_family="GC_rich"
 repeat_region 24565..24707
 /rpt_family="MIR"
 repeat_region 26013..26047
 /rpt_family="MIR2"
 repeat_region complement(27717..28532)
 /rpt_family="L1PB1"
 repeat_region complement(28553..28854)
 /rpt_family="AluSg"
 repeat_region complement(28860..29076)
 /rpt_family="L1A"
 repeat_region complement(29365..29561)
 /rpt_family="MIR"
 repeat_region complement(30050..30147)
 /rpt_family="MIR"

```

repeat_region      /rpt_family="AluJc"
                    complement(33629, .33732)
repeat_region      /rpt_family="MIR"
                    34434, .34729
repeat_region      /rpt_family="AluSx"
                    35070, .35361
repeat_region      /rpt_family="AluSx"
                    complement(36437, .36461)
repeat_region      /rpt_family="AT_rich"
                    complement(39967, .40261)
repeat_region      /rpt_family="AluSx"
                    complement(40741, .40915)
repeat_region      /rpt_family="AluSg"
                    40963, .41663
repeat_region      /rpt_family="L1PA2"
                    42042, .42352
repeat_region      /rpt_family="MLT1E"
                    42467, .42771
repeat_region      /rpt_family="AluSg"
                    42772, .42848
repeat_region      /rpt_family="L1MB8"
                    42849, .43151
repeat_region      /rpt_family="AluJb"
                    43152, .43318
repeat_region      /rpt_family="L1MA10"
                    43253, .43427
repeat_region      /rpt_family="L1MB8"
                    43429, .43672
repeat_region      /rpt_family="L1MB8"
                    complement(44188, .44481)
repeat_region      /rpt_family="AluSc"
                    44544, .45246
repeat_region      /rpt_family="L1TR8"
                    complement(46999, .47032)
repeat_region      /rpt_family="AT_rich"
                    complement(47411, .47431)
repeat_region      /rpt_family="AT_rich"
                    complement(47612, .47684)
repeat_region      /rpt_family="CATR1n"
                    complement(47946, .48239)
repeat_region      /rpt_family="AluJo"
                    complement(49827, .49872)
repeat_region      /rpt_family="AT_rich"
                    complement(50160, .50186)
repeat_region      /rpt_family="AT_rich"
                    50952, .51121
repeat_region      /rpt_family="MER5A"
                    complement(52140, .52284)
repeat_region      /rpt_family="AluJo"
                    53772, .53974
repeat_region      /rpt_family="MER1B"
                    54225, .54348
repeat_region      /rpt_family="MIR"
                    complement(54533, .54579)
repeat_region      /rpt_family="(TGA)n"
                    complement(54580, .54736)
repeat_region      /rpt_family="AluJb"
                    complement(54755, .54782)
repeat_region      /rpt_family="(CA)n"

```

```

Alignment Scores:
Pred. No.:          2,81e-08      Length:      106988
Score:             169.00         Matches:      30
Percent Similarity: 100.00%       Conservative: 0
Best Local Similarity: 100.00%    Mismatches:  0
Query Match:       100.00%        Indels:       0
DB:                9              Gaps:          0
JS-10-010-050a-2_COPY_1_30 (1-30) x ACC001226 (1-106988)

```

```

2y      1 MetAragtgglyalaglyalalargtgglyalargalaserTTPCySTTPAlaleuAlaleu 20
|||||

```

```

DB      23544 ATGGGGCGGGGGCGGGCGGGCTTCGGGAGCGCGCTTCCTGTGTGTGGCCCTGGCGCTG 23603
Qy      21 LeuTTPleuAlaValValProGlyTTPser 30
DB      23604 CTTGGCTCGCGGTGTTCGGGCTGTGCC 23633

RESULT 9
LOCUS    AL136440
DEFINITION Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.
ACCESSION AL136440
VERSION   AL136440.2 GI:10039473
KEYWORDS  HTG, HTGS_PHASE1, HTGS_CANCELLED.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS  Burton,J.
TITLE     Direct Submission
JOURNAL   Submitted (10-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
          requests: clonerequest@sanger.ac.uk
          On Sep 8, 2000 this sequence version replaced gi:6982057.

COMMENT
          Genome Center
          Center: Sanger Centre
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquery@sanger.ac.uk
          ----- Project Information
          Center project name: BA185124
          ----- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: plasmid; 108752; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Consensus quality: 164521 bases at least Q40
          Consensus quality: 16040 bases at least Q20
          Insert size: 164996; 1.7% error; agarose-fp
          Quality coverage: 4.08x in Q20 bases; sum-of-contigs Quality
          coverage: 4.16x in Q20 bases; agarose-fp

          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 14 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.

          1
          3401 contig of 3400 bp in length
          3500: gap of 100 bp
          3501 contig of 12944 bp in length
          16444: gap of 100 bp
          16445 contig of 2970 bp in length
          16545: gap of 100 bp
          16546 contig of 2627 bp in length
          19615: gap of 100 bp
          22241: gap of 100 bp
          22342 contig of 3185 bp in length
          25526: gap of 100 bp
          25627: gap of 100 bp
          25627 contig of 7444 bp in length
          33070: gap of 100 bp
          33170: gap of 100 bp
          33171 contig of 34450 bp in length
          67620: gap of 100 bp
          67720: gap of 100 bp
          67721 contig of 25693 bp in length
          93413: gap of 100 bp
          93414 contig of 20568 bp in length
          93514: gap of 100 bp
          114081: gap of 100 bp
          114181: gap of 100 bp
          126206: contig of 12025 bp in length
          126207: gap of 100 bp
          126307: gap of 100 bp
          144583: contig of 18277 bp in length
          144584: gap of 100 bp

```



```

/note="PRAM; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
3658..39964
/note="HS166D1"
4193..4497
/note="AluXx; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
4507..4804
/note="AluXx; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
4805..5105
/note="AluY; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
5174..5490
/note="AluXx; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
complement(10481..10623)
/note="MERS8; RepeatMasker predicted
23/10/2000"
/rpt_family="DNA/MER1 type"
complement(11229..11378)
/note="Aluud; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
complement(13829..14055)
/note="LIMB8; RepeatMasker predicted
23/10/2000"
/rpt_family="LINE/L1"
complement(14418..14787)
/note="LIMB8; RepeatMasker predicted
23/10/2000"
/rpt_family="LINE/L1"
complement(14808..14991)
/note="Aluud; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
complement(15301..15301)
/note="AluXx; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
15303..15414
/note="Aluud; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
16935..17110
/note="(CCTCG)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
17340..17414
/note="Aluud; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
join(19075..20523,21229..21437,21524..21707,21851..22074,
23338..23663,23748..23900,24287..24448,24537..24766,
24841..25007,25099..25277,25575..25752)
/gene="SOLH"
/note="PubMed:9722942"
/codon_start=1
/product="small optic lobes homolog"
/protein_id="AAK61233.1"
/db_xref="GI:1436701"
/db_xref="MIM:603267"
/db_xref="SPTREMBL:O75808"
/translation="MATVGMSCVRCFLNPAGROCSICAPRHKPDNLHILRLSVE

```

```

EOKMPCARCTFRNPLGKACEVCGFTPEPAPGAFLFVLNGVLPKPAALIGEPGSCQ
EAGPVATGIVATPEPARGQEDKDEBEKEOEEREGAAEPGGMACPRTLNTPVA
SSCGVGGPRSLIPRIIPALVYVEVVAAGVFPVAAAPPGLPBGAEANPATSQ
GPAEPVPPVPPSPSTLONNVPVRSRREVPOLOPPVPAAGSPSPAGCGAPQ
GSGWAGASRLABLSGRISLTVBERGCAACGCKTHGQEHBPBTPDGDADKSPC
SPSPDPTTWSCACTLNFPTVAPRGACCGCKTHGQEHBPBTPDGDADKSPC
GRCGVSASQAKARVLPERRGACACTLNLRAKHCAGCTPOLLVAORGAAP
LRRESEHVEORQTDGEAKALMENIVAFRENNVSPVDSPPGPSPVFPAGDSV
OQVRWMLRPOEINCSYFDRHATWSEFTLRPSDILGILGNCWFLSALVALERPD
LVERVMTSLCARGAYOVLCKDGTWTVTVDDMLPCDEAGCLFESQAKROLVAL
TERLATLHSYFLQAGRALBGLATLTGAPCESTALQSTNDREREDVDLIVAKM
LSSKEAPFLMGASCGGANNKVDVSAIYSLGIRPHNAYSILVRYVQGRRLRLNPMQ
RFSWNGSPSDSEWPMFPHLREGELMPHDSSEGVEMWEGDFVRYFDSDVICHSMDQ
ARVOCGSPSSAPAPGVATLTLVLRASLEFALPOGSRSDAVSHDLICILVERAT
FGSGHLSIGRLHNSKRAVKEFVSCDMLERPEAGVAVCAFNHMGPLGTIPAOAS
SPSGVPRASPERPCHTLAVYSRLVWVEPBAQPTLADAILLITESRGRHEGREG
MTCYTLTHMGAGLIVVENNRPKAVLHVQCCTSPNVVSRGSLRTODSVPLHROY
LVILISQDSGNAGFSITRHLARKAQAFLSMTWTSKGSHPPLTEVAGLHGRPL"
19443..19488
/note="(GGA)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
25280..25464
/note="C-rich; RepeatMasker predicted

Alignment Scores:
Pred. No.: 1,32e+03 Length: 256073
Score: 77.50 Matches: 15
Percent Similarity: 59.26% Conservative: 1
Best Local Similarity: 55.56% Mismatches: 10
Query Match: 45.86% Indels: 1
DB: 9 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x AB006464 (1-256073)
QY 4 GIYALAGLYAALALARY---GLYARGALASERTPYCTIPALALeulALeulTTP 22
Db 218086 GGGGCTGGGCTGGCCCGCAAGGGGCGGGGTGAGGTGGGCTGTGAGTCCGGTGG 218145
QY 23 LeuAlaValAlaPrGlyTTP 29
Db 218146 ACGGATGACGCCCGAGTTGG 218166

RESULT 15
AL137007 105779 bp DNA linear PRI 02-OCT-2000
LOCUS Human DNA sequence from clone RP1-30F19 on chromosome 6q12
DEFINITION Contains fragments of a gene similar to the Notch (Drosophila)
homolog, STS and GS8s, complete sequence.
ACCESSION AL137007
VERSION AL137007.9
KEYWORDS HTG; Notch.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 105779)
REFERENCE Williams,S.
AUTHORS Direct Submision
JOURNAL Submitted (01-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Jun 26, 2000 this sequence version replaced gi:5573795.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submision.
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the

```

sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Sw, SWISSPROT; Tr, TRMBL; Wp, WORMEP; Information on the WORMEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch6> RP1-303F19 is from the library RP1-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYRAC2 This sequence is the entire insert of clone RP1-303F19 The true left end of clone RP1-401P4 is at 105680 in this sequence.

FEATURES

source

```
1.105779
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="q12"
/clone="RP1-303F19"
/clone_lib="RP1-1"
843..1047
/feature="MER63A repeat: matches 7. .210 of consensus"
3043..3372
/feature="L1 repeat: matches 3905. .4233 of consensus"
3485..3623
/feature="L1 repeat: matches 4226. .4365 of consensus"
3624..3801
/feature="L1 repeat: matches 5971. .6146 of consensus"
3802..3815
/feature="L1 repeat: matches 4365. .4378 of consensus"
4413..4454
/feature="L1 repeat: matches 1. .365 of consensus"
5386..5777
/feature="L1 repeat: matches 1. .209 of consensus"
6538..6778
/feature="MER3 repeat: matches 1. .88 of consensus"
8132..8221
/feature="MER52A repeat: matches 1. .88 of consensus"
8222..8930
/feature="THB1A-internal repeat: matches 25. .737 of consensus"
8916..9626
/feature="THB1A-internal repeat: matches 867. .1580 of consensus"
9627..9979
/feature="THB1A repeat: matches 1. .354 of consensus"
10218..10903
/feature="match: GSS: Em:AQ321138"
10422..10793
/feature="match: GSS: Em:AQ113885"
10808..11106
/feature="AluB repeat: matches 7. .301 of consensus"
11321..11388
/feature="AluB repeat: matches 7. .301 of consensus"
12656..13022
/feature="MLT2CA repeat: matches 5. .460 of consensus"
13318..13318
/feature="match: GSS: Em:AQ817672"
13322..13322
/feature="match: GSS: Em:AQ450753"
12884..13584
/feature="match: GSS: Em:B83082"
12928..13479
/feature="match: GSS: Em:AQ003116"
13093..13135
/feature="match: GSS: Em:AQ003116"
15274..15509
/feature="MLT2CA repeat: matches 459. .505 of consensus"
```

```
repeat_region
/feature="MLT1E repeat: matches 310. .563 of consensus"
17051..17264
/feature="MLT1-INTERNAL repeat: matches 990. .1211 of consensus"
17349..17779
/feature="L1P13 repeat: matches 5736. .6156 of consensus"
17937..18338
/feature="MLT2B repeat: matches 1. .400 of consensus"
18369..18765
/feature="MLT1C repeat: matches 65. .462 of consensus"
18807..19243
/feature="match: GSS: Em:AQ034844"
19454..20541
/feature="L1P4 repeat: matches 5046. .6146 of consensus"
20712..20890
/feature="L1MC repeat: matches 1682. .1872 of consensus"
20964..21210
/feature="L1MC repeat: matches 1424. .1676 of consensus"
21293..22072
/feature="L1MC repeat: matches 5489. .6312 of consensus"
22102..22393
/feature="AluB repeat: matches 1. .293 of consensus"
22394..22951
/feature="L1MC2 repeat: matches 4769. .5328 of consensus"
22968..23066
/feature="L1M4 repeat: matches 2963. .3065 of consensus"
23070..24449
/feature="L1M47 repeat: matches 4601. .6003 of consensus"
24429..24982
/feature="L1 repeat: matches 3787. .4347 of consensus"
25033..25098
/feature="match: GSS: Em:AQ338604"
25120..25259
/feature="L1 repeat: matches 1. .74 of conserved"
25260..25603
/feature="L1 repeat: matches 4377. .4519 of consensus"
25305..25905
/feature="THB1B repeat: matches 1. .364 of consensus"
25604..25690
/feature="match: GSS: Em:AQ491381"
25806..25984
/feature="L1 repeat: matches 4518. .4604 of consensus"
25998..27419
/feature="L1MC repeat: matches 2278. .2099 of consensus"
27441..28042
/feature="MER52A repeat: matches 191. .1755 of consensus"
28570..29375
/feature="L1MC repeat: matches 1672. .2271 of consensus"
30178..30478
/feature="L1MC repeat: matches 944. .1803 of consensus"
30540..30562
/feature="AluSC repeat: matches 1. .303 of consensus"
30552..30838
/feature="Single clone region. Assembly confirmed by restriction digest"
31345..31761
/feature="AluSG repeat: matches 4. .294 of consensus"
33714..34154
/feature="L12 repeat: matches 1837. .2274 of consensus"
35163..35945
/feature="match: GSS: Em:AQ114098"
35947..36061
/feature="L12 repeat: matches 1837. .2666 of consensus"
36069..36886
/feature="L1M4 repeat: matches 5615. .5729 of consensus"
38551..38643
/feature="L1M4C repeat: matches 828. .1627 of consensus"
38845..39190
/feature="L12 repeat: matches 2631. .2731 of consensus"
39252..39633
/feature="THB1B repeat: matches 1. .364 of consensus"
/feature="L1M47 repeat: matches 5918. .6288 of consensus"
```

```

repeat_region      42264..42473
                    /note="I2 repeat: matches 2524..2745 of consensus"
repeat_region      42616..43264
                    /note="I2 repeat: matches 1594..2287 of consensus"
repeat_region      43381..43745
                    /note="I2 repeat: matches 1041..1394 of consensus"
repeat_region      43746..44031
                    /note="I2 repeat: matches 1..294 of consensus"
repeat_region      44032..44409
                    /note="I2 repeat: matches 629..1041 of consensus"
misc_feature       complement(44738..45299)
                    /note="match: GSS: Em:AQ081366"
misc_feature       45294..45604
                    /note="match: GSS: Em:AQ079097"
repeat_region      45491..45750
                    /note="HALI repeat: matches 1314..1585 of consensus"
misc_feature       complement(45588..46152)
                    /note="match: GSS: Em:B55710"
misc_feature       complement(45734..46193)
                    /note="match: GSS: Em:AQ224165"
repeat_region      46534..46589
                    /note="28 copies 2 mer at 80% conserved"
repeat_region      46593..46654
                    /note="31 copies 2 mer at 67% conserved"
repeat_region      47286..47467
                    /note="I2 repeat: matches 2555..2737 of consensus"

```

```

Alignment Scores:
Seq. No.:      864      Length:      105779
Score:          76.00    Matches:      14
Percent Similarity: 62.50%  Conservative: 1
Percent Local Similarity: 58.33%  Mismatches: 7
Every Match:      44.97%      Indels:      2
                    Gaps:      1

```

i-10-010-050A-2_COPY_1_30 (1-30) x AL137007 (1-105779)

```

      8 A1aArgG1ArgAlaSerTTPCyETTPAlaLeuAlaLeuTTPLeuAlaValVal--- 26
      |||||:::|||||||
      89610 GCAAGAGGTAAAGTTCATGGGTGGGACGCTGCCCTGTGGCTTTGCAGGGTACAG 89669
      27 ---ProGlyTTP 29
      |||||
      89670 CCCCCCTCCTGG 89681

```

Search completed: April 25, 2004, 03:25:56
 Job time : 652.083 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - nucleic search, using frame_plus_p2n model

on: April 24, 2004, 23:02:32 ; Search time 4011.01 Seconds
(without alignments)
2352.639 Million cell updates/sec

File: US-10-010-050a-2_COPY_31_346
Effect score: 1758
Sequence: 1 RVSGIPSRHRWVPIKRFDF.....IKTYEELPLIRNKLISGL 316

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

MODEL=frame+p2n.model -DEV=x1h
Q=/cpg2.1/USPTO.spool/US010050/rnat.22042004.113204.27556/app.query.fasta_1.1372
DB=EST -QFMT=faa2ap -SUFFIX=rst -MINMATCH=0.1 -LOOPEC=0 -LOOPEXT=0
UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
LOCALIZATION=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US010050 @CGN 1.1 6283 @rnat.22042004.113204.27556 -NCPU=6 -ICPU=3
NO_MMAP -LARGEIODRV -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: EST:
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mus:*
23: em_gss_mam:*
24: em_gss_pro:*
25: em_gss_rnd:*
26: em_gss_phg:*
27: em_gss_vit:*
28: gb_gss1:*

29: gb_gss2:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1570	89.3	1201	9	AL546472	AL546472 AL546472
2	1423	80.9	2356	11	AK043247	AK043247 Mus muscu
3	1418	80.7	2247	11	AK032293	AK032293 Mus muscu
4	1415	80.5	2433	11	AK085741	AK085741 Mus muscu
5	1237	70.4	689	13	B0621797	B0621797 UI-H-FL1
6	1225.5	69.7	924	14	CA488543	CA488543 AGENCOURT
7	1225	69.7	923	13	BX331615	BX331615 BX331615
8	1214	69.1	1201	9	AL571805	AL571805 AL571805
9	1211	68.9	931	10	B8873363	B8873363 601450425
10	1191	67.7	869	12	B1090566	B1090566 602855673
11	1167	66.4	908	13	BX370641	BX370641 BX370641
12	1159	65.9	974	13	BQ919385	BQ919385 AGENCOURT
13	1158	65.9	915	12	B1080303	B1080303 602876736
14	1143	65.0	787	12	B1223533	B1223533 602941923
15	1136	64.6	641	14	CB215456	CB215456 NISC_np06
16	1124.5	64.0	1077	13	BX342662	BX342662 BX342662
17	1121	63.8	938	13	BQ926170	BQ926170 AGENCOURT
18	1113	63.3	755	13	BQ443453	BQ443453 UI-M-EVO-
19	1076	61.2	1098	10	BF982158	BF982158 602308976
20	1071	60.9	777	10	AW475730	AW475730 UNF1C09.Y
21	1069	60.8	822	12	B1734549	B1734549 602919530
22	1058.5	60.2	858	12	B1156421	B1156421 602919530
23	1041.5	59.2	853	12	B1456284	B1456284 603172673
24	1028	58.5	915	13	CK128221	CK128221 AGENCOURT
25	1018.5	57.9	915	13	BX757396	BX757396 BX757396
26	1012	57.6	564	14	CB297292	CB297292 12B22045
27	969.5	55.1	676	13	BX57383	BX57383 BX757383
28	959	54.6	676	12	BG820113	BG820113 602782325
29	953	54.2	680	14	CF724176	CF724176 UI-M-GZO-
30	952	54.2	688	14	CF593022	CF593022 AGENCOURT
31	938.5	53.4	766	12	B182240	B182240 UNTL-P-FN-
32	931	53.0	767	9	AJ451053	AJ451053 AJ451053
33	922	52.4	869	14	CF378929	CF378929 AGENCOURT
34	920.5	52.4	854	12	B1082103	B1082103 602877265
35	901	51.3	607	12	BM849262	BM849262 K-EST0129
36	899	51.1	607	28	AQ309525	AQ309525 CITBT-EI-
37	891.5	50.7	922	14	CA789305	CA789305 AGENCOURT
38	882	50.2	661	13	BY734042	BY734042 BY734042
39	880	50.1	516	14	CB455600	CB455600 712753 MA
40	857.5	48.8	736	12	B1077452	B1077452 602872262
41	856	48.7	720	12	B1111605	B1111605 602895967
42	843	48.0	623	13	BQ417448	BQ417448 IK8505.Y
43	841	47.8	604	10	BB622627	BB622627 BB622627
44	841	47.8	615	9	A1746877	A1746877 u108612.Y
45	834.5	47.5	706	10	BF739857	BF739857 7039A03.X

ALIGNMENTS

RESULT 1
AL546472
LOCUS AL546472 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
DEFINITION Clone GSD01030Y01 5-PRIME, mRNA sequence.
ACCESSION AL546472
VERSION AL546472.2 GI:31268306
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1201)

AUTHORS
TITLE
JOURNAL
COMMENT

Li, M. B., Gruber, C., Jesse, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12879620.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1030CE01QP1&cluster=7238.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue genoscope sequence ID : CS0D1030CE01QP1.
Location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1030Y01"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-cligo (dt)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMSPORT 6 vector. Library was normalized."

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 4,01e-164 Length: 1201
Score: 1570.00 Matches: 292
Percent Similarity: 96.37% Conservative: 0
Best Local Similarity: 96.37% Mismatches: 10
Query Match: 89.31% Indels: 3
DB: 9 Gaps: 1

US-10-010-050A-2_COPY_31_346 (1-316) X AL546472 (1-1201)

QY 1 ArgValSerGlyTleProSerArgHisTTPProValProTyrTlyAspPhe 20
DB 153 CGGGTCCGGGATCCCTCCCGGCGCCACTGCGCGTCCCTACAGGCTTGACCTC 212
QY 21 ArgProTyrProAspProTyrCysGlnAlaTyrTyrPheCysProThrGlySerPro 40
DB 213 CGTCCAAACCTGATCTTATGTCAAGCTTAAGTACTTTCTGCCAATGGCTCACC 272
QY 41 TleProValMetGluGlyAspAspAplleGluValPheArgGluGlnAlaProValTTP 60
DB 273 ATCCGAGTTATGAGGGGTGATGATGACATTGAAGTTTTCGATTACAGCCCAATG 332
QY 61 GluPheTyrTyrGlyAspLeuLeuGlyHisLeuValIleMetHisAspAlaIleGlyPhe 80
DB 333 GAATTTAAATATGAGACCTCCCGGAGCACTTGAAATATGATGATATGATGATGATG 392
QY 81 ArgSerThrLeuThrGlyAspTyrTyrMetGluTyrTyrGluLeuPheGlnLeuGly 100
DB 393 AGAAGTACATTAATCTGGCAAGAACTACACAAATGAAATGAAATCTTTTCCAACTGGC 452
QY 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
DB 453 AATGTACATTTCCCATCTCCGACCTCGAAGTGAATGAGTCCCTTCTGGTATATCAAGC 512
QY 121 AlaAlaCysPhePheGluGlyIleAspAspValHisTyrTyrGluGlnLeuVal 140
DB 513 GCTGCCGCTTTTGGGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 572
QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaTyrTyrValIleGlnAsp 160
DB 573 CAAAGTCAACCTATATAGGAACATTTTCAACCAATGGCAAGTGGGGAACAGAC 632
QY 161 AsnGluThrGlyIleTyrTyrGluThrTyrPheValIleAspProGluTyrGlyAla 180

DB 633 AATGAACGGAATTTATTTATGACATGATGATGATGATGATGATGATGATGATGATG 692
QY 181 GluThrTyrPheAspSerTyrAspCysSerTyrSphValLeuArgThrPheAsnLeu 200
DB 693 GAGACATGTTGATTCCTACGACTGCTTCCAAATTTGTGAAGCCCTTTAACAGTTG 752
QY 201 AlaGluPheGlyAlaGluPheAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220
DB 753 GCTGAATTTGGAGCAGATGTCAGAACATGAAACCACTATACAAATATTTCTTTAC 812
QY 221 SerGlyGluProThrTyrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsn 240
DB 813 AGTGGAGAACCTTACTTATCTGGAAATGAAACATCTGTTTTCGGGCAACAGAAACAG 872
QY 241 ThrLeuGlyLeuAlaIleLeuAspGlyTyrTyrProPheAspProHisLeuProThrTyr 260
DB 873 ACCTTGTTTACCAATTAATAATTTATTTTACCCCTTCAACACATTTGCCAATCTAA 932
QY 261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisGlyGlnPheTyr 280
DB 933 GAATTTCTGTTGATCTTCTGCAAAATTTGATGATGATGATGATGATGATGATGATG 992
QY 281 LeuPheTyrAsnPheGluTyrTyrPheLeuProMetTyrSphProPheIleTyr 300
DB 993 TTGTTTATTAATTTGGA-TATGGTTTMM---CYATGAAMTCCCTTATTAAMA-ACA 1047
QY 301 TyrGluGlu 303
DB 1048 TATGAGGA 1056

RESULT 2
AK043247
LOCUS
DEFINITION
MUS MUSCULUS 7 DAYS NEONATE CEREBELLUM CDNA, RIKEN FULL-LENGTH
ENRICHED LIBRARY, CLONE: A730075N08 PRODUCT: SIMILAR TO
CERIOD-LIPOPOSCINOSIS NEURONAL PROTEIN 5 (CLIN5 PROTEIN) [Homo
sapiens], full insert sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL
MEDLINE
PUBMED
REFERENCE
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE

AK043247 2356 bp mRNA linear HTC 19-SEP-2003
AK043247.1 GI:26089577
MUS MUSCULUS (house mouse)
MUS MUSCULUS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
2049374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Katsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL REFERENCE Nature 409, 685-690 (2001)

AUTHORS 5

TITLE The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

JOURNAL REFERENCE Analyses of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

AUTHORS Nature 420, 563-573 (2002)

TITLE 6 (bases 1 to 2356)

JOURNAL REFERENCE Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, T., Kojima, Y., Kondo, S., Komori, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku, A., Takahashi, S., Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL REFERENCE Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Please visit our web site for further details.

SOURCE URL: http://genome.gsc.riken.go.jp/

LOCATION URL: http://fantom.gsc.riken.go.jp/

LOCATION Location/Qualifiers

FEATURES 1. 2356

SOURCE /organism="Mus musculus"

FEATURES /mol_type="mRNA"

FEATURES /strain="C57BL/6J"

FEATURES /db_xref="FANTOM DB:A730075N08"

FEATURES /db_xref="MGI:2408405"

FEATURES /db_xref="taxon:10090"

FEATURES /clone="A730075N08"

FEATURES /tissue_type="cerebellum"

FEATURES /clone_lib="RIKEN full-length enriched mouse cDNA library"

FEATURES /dev_stage="7 days neonate"

FEATURES 1. 2356

FEATURES /note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CIN5 PROTEIN) [Homo sapiens] (SWISSPROT:O75503, evidence: FASTA, 73.9%ID, 86.7%length, match=1041)"

ALIGNMENT SCORES:

RED. NO.: 2,28e-147

SCORE: 1423.00

PERCENT SIMILARITY: 89.03%

FAST LOCAL SIMILARITY: 79.35%

NESTED MATCH: 80.94%

INDELS: 11

GAPS: 0

US-10-010-050a-2_COPY_31_346 (1-316) X AK03247 (1-2356)

Y 7 SetAArgArgHisTrpProValProTyrTyrGlyArgPheAspPheArgProLysProAspPro 26

b 148 TCTGGGCAAGCGTGGCGGCGCTTCTCTTCCGTCGGAAGACAGATCC 207

Y 27 TycGagGlnAlaValGlyTyrThrPheCysProThrGlySerProIleProValMetGluGly 46

DB 208 TACTGTCAAGCTAAGTATTAATTCTGTCTACCGGCTGCGCCATCCAGTTATGAGAGAC 267

QY 47 AspaSpaepIleGluValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAsp 66

DB 268 AATGACGTCATCGAGGCTTACGACTACCAACCCCGAATTTGAGATTTAAATATGAGAAC 327

QY 67 LeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGly 86

DB 328 CTCCTGGACACTTTAAACTTATGTCATACACCCGCGGAGATTCAGAGACACATCGACAGC 387

QY 87 LysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHis 106

DB 388 AAGAACTACACATTTAGATGATGATGATGATTTCCAGCTGGGCAACTGATCAATTTCCAC 447

QY 107 LeuArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGlu 126

DB 448 CTCGGCGCTCGCAAGAGCGCTCCCTTCGCGGTATACCAAGGGGCGACCTGCTTTTGTGA 507

QY 127 GlyIleAspAspValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSer 146

DB 508 GGAATGATGATTAACACTGGAAGAAAACGGACACTGTCAGTCTGTCACATATCC 567

QY 147 GlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyr 166

DB 568 GGAACACATTTTAAACAAGTGGCCGAGTGGGGAAGCAGACATGAACTGGCATTTAT 627

QY 167 TyrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSer 186

DB 628 TATGAGACATGACCGTCCGACCGCCGACGAGCAAGGGGCGACAGCTGTTGAGTCC 687

QY 187 TyrAspCysSerLysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGlu 206

DB 688 TACGACTGTTCGAATTTTGTCTTAAAGACATTAAGAAATTTGGCTGAATTTGGAACGAA 747

QY 207 PheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyr 226

DB 748 TTCAGAAGATGAAACAACTATACGAAATATTTCTTACAGTGGAGAGCTATTATAC 807

QY 227 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle 246

DB 808 CTGGGAAATGAAACATCTATTTTGGGCCCAAGGAAACAAACCTTCTTGGCCATA 867

QY 247 LysArgPheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuLeuSerLeu 266

DB 868 AAAAATTTTATAGCCCTTCGACCGATTTGTCAACCAAGATTTCTGATGATTTTC 927

QY 267 LeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGlu 286

DB 928 TTGAATAATTTTGAATACAGTATTAACACAGACAGTCTTACTGTTTATAACTTTGAG 987

QY 287 TyrTrpPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeu 306

DB 988 TATGTGTTTCTACATGAAACCCCTTTCGCAAAATACATACGAAAGAACCCCGTTA 1047

QY 307 ProIleArgAsnLysThrLeuSerGlyLeu 316

DB 1048 CCAACCCGACATACACATTTTACCACTTG 1077

RESULT 3

AK032293 2247 bp mRNA linear HTC 18-SEP-2003

LOCUS AK032293

DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430516P20 product:similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CIN5 PROTEIN) [Homo sapiens], full insert sequence.

ACCESSION AK032293

VERSION AK032293.1 GI:26328112

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
AUTHORS
2 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Itoh, A., Yamamoto, R., Matsuno, H., Sakaguchi, S., Ikegami, T., Kasahara, K., Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE
AUTHORS
3 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
MEDLINE
PUBMED
REFERENCE
AUTHORS
4 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE
PUBMED
REFERENCE
AUTHORS
5 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, F., Fukuda, S., Furuno, M., Hasegawa, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirose, T., Hori, F., Imoto, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, K., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Science Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscl.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Science Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
FEATURES
source
1. .2247
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:6430516P20"
/db_xref="WGI:2335762"

ORIGIN
Alignment Scores:
Pred. No.: 7,77e-147 Length: 2247
Score: 1418.00 Matches: 245
Percent Similarity: 88.718 Conservative: 30
Best Local Similarity: 79.038 Mismatches: 35
Query Match: 80.664 Indels: 0
DB: 11 Gaps: 0
US-10-010-050a-2_copy_31_346 (1-316) x AK032293 (1-2247)
QY 7 SerArgArgHisIstProValProIlytYSAgPheAspPheArgProIysProAspPro 26
Db ::::
QY 39 TCtGGGAGCAGCTGCGCGGCGCATCAAGCACCCTCTCTCTCGTCCGAGACAGANTCC 98
Db TCtGGGAGCAGCTGCGCGGCGCATCAAGCACCCTCTCTCTCGTCCGAGACAGANTCC 98
QY 27 TyTCyGGLAlAlaYstYrThrPheCySProThGlySerProIleProValMetGluGly 46
Db TACTGTCAAGCTAAGTAACTTCTGTGCTTACCGGCTCGCCATCCAGTTATGAAGAC 158
QY 99 TACTGTCAAGCTAAGTAACTTCTGTGCTTACCGGCTCGCCATCCAGTTATGAAGAC 158
Db TACTGTCAAGCTAAGTAACTTCTGTGCTTACCGGCTCGCCATCCAGTTATGAAGAC 158
QY 47 AspAspAspIleGluValPheArgLeuGlnAlaProValTyrGluPheIysTyrGlyAsp 66
Db ::::
QY 159 AATGACGCTATCGACAGCTTACGACATCAAGCCCGATTGGGAATTAATATGAGAC 218
Db AATGACGCTATCGACAGCTTACGACATCAAGCCCGATTGGGAATTAATATGAGAC 218
QY 67 LeuLeuGlyHisLeuIysIleMetHisAspAlaIleGlyPheArgSerThrLeuThGly 86
Db ::::
QY 219 CTCCTGGGACACTTAACTTATGATGATGACGCGGTGGATTGAGACACATGACAGGC 278
Db CTCCTGGGACACTTAACTTATGATGATGACGCGGTGGATTGAGACACATGACAGGC 278
QY 87 LysAsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCyThrPheProHis 106
Db ::::
QY 279 AAGAACTACACATTAAGTGTGATGAACCTTTCCAGCTGGGCAATGTATCCCCAC 338
Db AAGAACTACACATTAAGTGTGATGAACCTTTCCAGCTGGGCAATGTATCCCCAC 338
QY 107 LeuArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGlu 126
Db ::::
QY 339 CTCGGGCTTACCAAGACGCTCCCTTCTGTGTGAACCAAGGCGCAGCTGTTTGGAA 398
Db CTCGGGCTTACCAAGACGCTCCCTTCTGTGTGAACCAAGGCGCAGCTGTTTGGAA 398
QY 127 GlyIleLeuAspAspValHisTyrPlyGluAsnGlyThrLeuValGlnValAlaThrIleSer 146
Db ::::
QY 399 GGAATAGATATATAAAGCTGGAAGAAACGGGACATGCTGATGGAACCATATCC 458
Db GGAATAGATATATAAAGCTGGAAGAAACGGGACATGCTGATGGAACCATATCC 458
QY 147 GlyAsnMetPheAsnGlnMetAlaIleTyrValIysGlnAspAsnGluThrGlyIleTyr 166
Db ::::
QY 459 GGAACACATTTTAAACAAGTGGCCGAGTGGTGAAGAGCAATGAATCGGATTTAT 518
Db GGAACACATTTTAAACAAGTGGCCGAGTGGTGAAGAGCAATGAATCGGATTTAT 518
QY 167 TyGluThrTyrAsnValIysAlaSerProGluIysGlyAlaGluThrTyrPheAspSer 186
Db ::::
QY 519 TATAGACATGAGAGCGTCCGAGCGGCGCCAGCAAGAGGCGCCAGCGTGTGAGAGTCC 578
Db TATAGACATGAGAGCGTCCGAGCGGCGCCAGCAAGAGGCGCCAGCGTGTGAGAGTCC 578
QY 187 TyTAspCySerSerIysPheValLeuArgTyrPheAsnIysLeuAlaGluPheGlyAlaGlu 206
Db ::::
QY 579 TACACAGTTCGAATTTTGTCTTAAGACATATGAATATGCGTGAATTTGGAACGAA 638
Db TACACAGTTCGAATTTTGTCTTAAGACATATGAATATGCGTGAATTTGGAACGAA 638
QY 207 PheIysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyr 226
Db PheIysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyr 226

```

|||||
639 TTCAAGAGATAGAAACAACTATACGAAATATTTCTTTCAGTGGAGACCTATTATAC 698
227 LeuGlyAangUuThrSerValPheGlyProthrglyAenUsthrLeuGlyLeuAlaIle 246
699 CTGGGAATGAAACCTCATTTATTTTGGGCCCAAGAAACAAAGACTTGTGCTTGGCCATA 758
247 LysArgPheTyTyTyProPheIysProHisIleuProthrglyGluPheLeuIleuSerIleu 266
759 AAAAATTTTATTTGGCCCTTACGACCGATTTGTCAACCAAGATTTTCTGATGAAATTC 818
267 LeuGlnIlePheAspAlaValIleValHisIySegInPheTyTyLeuPheTyAspPheGlu 286
819 TTGAATAATTTTGTATACAGTGTATATACACAGACAGTCTCTACTGTTTATTAACCTTGAG 876
287 TyTTPheIleuProMetIysPheProPheIleTyIleTyTyGluGluIleProIleu 306
879 TATTTGTTTCTTACCATGAAACCCCTTTGTCAAAATTAACATGACGAAGAACCCCGTTA 936
307 ProIleArgAsnTyThrIleuSerGlyLeu 316
939 CCTACCCACATACAAACATTTTACGACTTG 968

RESULT 4
1085741 2433 bp mRNA linear HTC 20-SEP-2003
XUS Mus musculus 10 days lactation, adult female mammary gland cDNA,
PRINTITION RIKEN full-length enriched library, clone:D730033P03
PRODUCT: similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5
PROTEIN) [Homo sapiens], full insert sequence.
SESSION AK085741
SESSION AK085741.1 GI:26102938
XUS HTC; CAP trapper.
XUS Mus musculus (house mouse)
XUS Mus musculus
XUS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ORGANISM

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
2 Carninci, P. and Hayashizaki, Y.
3 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kikunishi, T., Tachino, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Hatake, A.,
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Nature Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
TITLE

```

```

JOURNAL of 60,770 full-length cDNAs
REFERENCE Nature 420, 563-573 (2002)
AUTHORS 6 (bases 1 to 2433)
Aachi, J., Aizawa, K., Akimura, T., Arakawa, T., Boro, H., Carninci, P.,
Fukuda, S., Furuno, M., Hamagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imctani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Onno, M., Ohnito, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
FEATURES
source
1. 2433
location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:D730033P03"
/db_xref="MG1:2422937"
/db_xref="taxon:10090"
/clone="D730033P03"
/sex="female"
/tissue_type="mammary gland"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days lactation, adult"
1. 2433
/note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5
(CLN5 PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence:
FASTA, 73.3%ID, 86.7%length, match=1041)"
ORIGIN
misc_feature
1. 85e-146
Alignment Scores: 2433
Pred. No.: 1415.00 Matches: 245
Score: 88.71% Conservative: 30
Percent Similarity: 79.03% Mismatches: 35
Best Local Similarity: 80.43% Indels: 0
Query Match: 11 Gaps: 0
DB:
US-10-010-050A-2_COPY_31_346 (1-316) x AK085741 (1-2433)
QY 7 sepArgGhGtTTPProValProTyTyLysArgPheAspPheArgProLysProAspPro 26
|||||
DB 225 TCTGGGACGCGTGGCCGCTGCTTCAACGCGTCTCTTTCGCTCGAAGACAGATCC 284
QY 27 TyTyGlnAlaLysTyTyThrPheTyProthrglySerProIleProValMetGluGly 46
|||||
DB 285 TACTGCAAGCTAAGATATCTTCTGCTTCAACGCGCTCGCCATCCAGTTATGAAGAC 344
QY 47 AspAspAspIleIleValPheArgLeuGlnAlaProValTTPgluPheTyTyGlyAsp 66
|||||
DB 345 AATGAGGTGATTCAGGCTTACGACTACAAAGCCCAATTTGGAATTTAATATGAGAC 404
|||||
QY 67 LeuLeuGlyHisIleuTyIleMetHisAspAlaIleGlyPheArgSerThrIleuThrgly 86

```

```

|||||
b 405 CTCTGGGACACTTTAACTTATGACATGACCGCGTGGGATTTCAGGACACACGACGAC 464
y 87 LysAsnTyrThrMetGluTrpTyrGluLeuPheGluLeuGlyAsnGlyThrPheProHis 106
b 465 AAGAACTACACAAATGATGATGATGAACCTTTCCAGCTGGGCACTGTACATTTCCCAAC 524
y 107 LeuArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGlu 126
b 525 CTCCGGCTTACCAAGACCGCTCCCTCTGTGTATGACCAAGGGGACGCTCTTTTGA 584
y 127 GlyLeuAspValHisTrpTyrGluAsnGlyThrLeuValGlnValAlaThrIleSer 146
b 585 GCAATATGATGATTAACACTGGAAGAAAACGGGACACTGTCACTGTTCACCAACATATCC 644
y 147 GlyAsnMetPheAsnGlnMetAlaLeuTrpValGlnGlnAsnGluThrGlyIleTyr 166
b 645 GGAACACATTTCACAAAGTGGCGAGTGGGTGAAGACAGACATGAATCTGGGATTTAT 704
y 167 TGTGTTTTPAenValIleValAserProGluGlyIleGluThrTrpPheAspSer 186
b 705 TATAGACATGACGCGTCCGACCGGCTCAGGACAAAGGGCCCAAGCTGTGTTCAAGTCC 764
y 187 TyrAspCysSerIysPheValLeuArgTrpPheAsnIleValGluAlaGluThr 206
b 765 TAGACCTGTTGATTTTGTCTTAAAGACATATTAAGAAATTTGGCTGATTTGGAAACGAA 824
y 207 PheIysAsnIleGluThrAsnTyrThrArgIlePheLeuTrpSerGlyGluProThrTyr 226
b 825 TTCAAGAGAAGAAACAACTATACGAAATATTTCTTACAGTGGAGAGCCATTTTAC 864
y 227 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIleValGluIle 246
b 885 CTGGGAATGAAACATCTATTTTGGGCGCCAAAGAAACAAAGACTCTTGCTTGGCATA 944
y 247 LysArgPheTyrTyrProPheIysProHisLeuProThrIysGluPheLeuSerLeu 266
b 945 AAAAATTTTATGGCCCTTCAGACCGGATTTGCCAACCAAGATTTCTGATTAATTTTC 1004
y 267 LeuGlnIlePheAspAlaValIleValHisIysGlnPheTyrLeuPheTyrAsnPheGlu 286
b 1005 TTGAAATATTTTGTATACGATATTAACACAGACAGTCTTCTGTTTATTAACCTTGAG 1064
y 287 TyrTrpPheLeuProMetIysPheProPheIleIysIleThrTyrGluGluIleProleu 306
b 1065 TATGCTTTCTTACCAATGAACACCCCTTGTCAAAATTAATACGAAGAAACCCCGTAA 1124
y 307 ProIleArgAsnIleTrpLeuSerGlyLeu 316
b 1125 CCTACCCGACATACAAACATTTACCGACTTG 1154

```

RESULT 5
 BUE21797/c 689 bp mRNA linear EST 23-SEP-2002
 LOCUS UI-H-FL1-bgb-i-07-0-UI-s1 NCI CGAP_FIL Homo sapiens cDNA clone
 DEFINITION UI-H-FL1-bgb-i-07-0-UI 3', mRNA sequence.
 BUE21797
 ACCESSION BUE21797.1 GI:23288012
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCI CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 Tissue Procurement: James Martin
 Tissue Procurement: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

JOURNAL COMMENT
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: James Martin
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 Seq primer: ML3 FORWARD
 POLYA=yes.
 FEATURES
 source
 Location/Qualifiers
 1. 689
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-FL1-bgb-i-07-0-UI"
 /tissue_type="Cell lines"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_id="NCI CGAP_FIL"
 /note="Organ: Chondrosarcoma; Vector: pT73-Pac
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA,
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pT73-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 GAGGTCGGTG. The cell lines were provided by Dr. James
 Martin from the University of Iowa.
 TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix
 TAG_LIB=UI-H-FL1
 TAG_SEQ=GAGGTCGGTG"

ORIGIN
 Alignment Scores:
 Pred. No.: 2 71e-127 Length: 689
 Score: 1237.00 Matches: 222
 Percent Similarity: 100.00% Conservative: 3
 Best Local Similarity: 98.67% Mismatches: 0
 Query Match: 70.36% Indels: 0
 DB: 13 Gaps: 0

US-10-010-050a-2_copy_31_346 (1-316) x BUE21797 (1-689)
 Oy 79 GlyPheArgSerThrLeuThrGlyIleValAsnTyrThrMetGluTrpTyrGluLeuPheGlu 98
 Db 688 GGAATTCAGAAATGATTAATGCGGCAAGACATGCAATGAATGATGAACTTTCCAA 629
 Oy 99 LeuGlyAsnCysTrpPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsn 118
 Db 628 CTGGCAACTGTATACATTTCCCATCTCCGACCTGAATGAATGGATGCCCTTTCTGTGTAT 569
 Oy 119 GlnGlyAlaAlaCysPhePheGlnGlyIleAspAspValHisTrpTyrGluAsnGlyThr 138
 Db 568 CAAAGCGCTGCTCTCTTTTGGAGGAATGATGATGTCACTGGAAGAAATGGGACCA 509
 Oy 139 LeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleValTyrValIle 158
 Db 508 TTAGTTCAAGTAGACATATATACAGAAACATGTTCAACCAATGGCAAAAGTGGGTGAA 449
 Oy 159 GlnAspAsnGluThrGlyIleTyrTyrGluThrTrpAsnValIleValAserProGluIle 178
 Db 448 CAGACCAATGAAGAAAGAAATTTTGGAGCAATGAAATGTAAGCCGACCAAGAAAG 389
 Oy 179 GlyValGluThrTrpPheAspSerTyrAspCysSerIysPheValLeuArgTrpPheAsn 198
 Db 388 GGGGCAAGACATGGTTGATTCCTACAGACGTTCCAATTTGTGTAAAGACCTTTTAC 329
 Oy 199 LysLeuAlaGluPheGlyAlaGluPheIysAsnIleGluThrAsnTyrThrArgIlePhe 218
 Db 328 AAGTGGCTGAATTTGGAGCAAGTTCGAAGAACATGAACCACTATACAAATATTT 269

219 LeuTySerGlyGluProThrTyrLeuGlyAAsnGluThrSerValPheGlyProThrGly 2:8
 268 CTTTACAGTGGAGAACTACTATTGGGAATGAACATCTGTTTGGCCAAACAGA 209
 239 AsnTyThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuPro 2:58
 208 AACAAAGACTCTGGTTGGTCCATATAAAGATTATTATACCCCTCAACACCATTTGGCA 149
 259 ThrTyGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisGlyGln 2:78
 148 ACTAAGAGATTTCGTAGTCTCTGCAAAATTTTGATGAGATGATTTGCACAAACAG 89
 279 PheTyrLeuPheTyrAsnPheGluTyrTyrTrpPheLeuProMetLysPheProPheIleLys 2:98
 88 TTTATTTGTTTATATATTGAAATATGTTTACTATGAATATCCCTTTATATAA 23
 299 IleThrTyrGluGln 303
 28 ATTAACATTGAAAGA 14

RESULT 6
 CA488543
 LOCUS
 DEFINITION AGENCOURT 10808884 MAPcl Homo sapiens cDNA clone IMAGE:6720309 5',
 mRNA sequence.
 ACCESSION CA488543
 VERSION CA488543.1 GI:24950702
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE
 1 (baee 1 to 924)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM14279 row: j column: 21
 High quality sequence stop: 637.

Location/Qualifiers

1..924

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6720309"

/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-HMEL, LNCaP"

/lab_host="EMDHL08"

/clone_1ib="MAPcl"

/note="Vector: PCMV-SPORE6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkok Lee & Ira Pastan. Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:

Pred. No.: 7,48e-126 Length: 924
 Score: 1225.50 Matches: 229
 Percent Similarity: 87.73% Conservative: 7
 Best Local Similarity: 85.13% Mismatches: 20

Query Match: 69.71% Indels: 13
 DB: 14 Gaps: 3
 US-10-010-050a-2_copy_31_346 (1-316) x CA488543 (1-924)

QY 1 ArgValSerGlyIleProSerArgArgHisIleTrpProValProTyrIlyArgPheAspPhe 20
 106 CGAGTCTCGGAGCATCCCTCCGGCGGCACCTGGCGGCTGCCTACAGCGCTTTGACATTC 165
 DB 106 CGAGTCTCGGAGCATCCCTCCGGCGGCACCTGGCGGCTGCCTACAGCGCTTTGACATTC 165
 QY 21 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40
 166 CGTCAAAACCTGATCCTTATGTCAGACTAAGTATCTTCTGTCCAACTGGCTCACCT 225
 DB 166 CGTCAAAACCTGATCCTTATGTCAGACTAAGTATCTTCTGTCCAACTGGCTCACCT 225
 QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrp 60
 226 ATCCAGTTATGAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 285
 DB 226 ATCCAGTTATGAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 285
 QY 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe 80
 286 GAATTTAATATGAGAGACTCTCTGGACACTTGAATATATGATGATGATGATGATGATGATG 345
 DB 286 GAATTTAATATGAGAGACTCTCTGGACACTTGAATATATGATGATGATGATGATGATGATG 345
 QY 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrGlyLeuPheGluLeuGly 100
 346 AGAAGTACATTAACTGGCAGAGAACTACCAATGAAATGATGATGATGATGATGATGATGATG 405
 DB 346 AGAAGTACATTAACTGGCAGAGAACTACCAATGAAATGATGATGATGATGATGATGATGATG 405
 QY 101 AsnCyThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
 406 AACTGTACATTTCCCACTCCGACTCCGACTGAATGATGATGATGATGATGATGATGATGATG 465
 DB 406 AACTGTACATTTCCCACTCCGACTCCGACTGAATGATGATGATGATGATGATGATGATGATG 465
 QY 121 AlaAlaCysPhePheGluGlyIleAspAspValHisIleTrpLysGluAsnGlyThrLeuVal 140
 466 GCTGCTCTCTTTTGTGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 525
 DB 466 GCTGCTCTCTTTTGTGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 525
 QY 141 GluValAlaThrIleSerGlyAsnMetLeuGlnMetAlaLysTrpValLysGlnAsp 160
 526 CAAGTAGACATATATGAGAAACATGTTCAACCAAAAGGCAATGGGTGAAACAGGAGC 585
 DB 526 CAAGTAGACATATATGAGAAACATGTTCAACCAAAAGGCAATGGGTGAAACAGGAGC 585
 QY 161 AsnGluThrGlyIleTyrTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180
 586 AATGAAACAGCAATTTATATGAGACATGAGATGTAAGGCAAGCCAGCAAGAAAGGGGGCA 645
 DB 586 AATGAAACAGCAATTTATATGAGACATGAGATGTAAGGCAAGCCAGCAAGAAAGGGGGCA 645
 QY 181 GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200
 646 GAGACATGTTGATTCCTACAGACCTGTCCTCAAAATTTGCTTAAAGACCTTTAAACAGTTG 705
 DB 646 GAGACATGTTGATTCCTACAGACCTGTCCTCAAAATTTGCTTAAAGACCTTTAAACAGTTG 705
 QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArg-IlePheLeuT 220
 706 GCTGAATTTGAGCAGAGTTCAGAGAACTAGAAACCAATATACAGAAATATTTCTTT 765
 DB 706 GCTGAATTTGAGCAGAGTTCAGAGAACTAGAAACCAATATACAGAAATATTTCTTT 765
 QY 220 YrSerGlyGluPro---ThrTyrLeuGlyAsnGluThrSer-----ValPheGly---- 235
 766 ACAGTGGGAGAAACCTTACTTATCCGGGGAATTTGAAAACATTCGATTTTGGGGCCT 825
 DB 766 ACAGTGGGAGAAACCTTACTTATCCGGGGAATTTGAAAACATTCGATTTTGGGGCCT 825
 QY 236 -----ProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgP 249
 826 AACAAAGAAACCAAGAACTCCTTGGGTTCATACCCCATTAACAAAAAATTTTAAAT 885
 DB 826 AACAAAGAAACCAAGAACTCCTTGGGTTCATACCCCATTAACAAAAAATTTTAAAT 885
 QY 249 heTyrTyrProPheLysProHis 256
 886 TAACCGCCCTTCTCAAAACAC 908
 DB 886 TAACCGCCCTTCTCAAAACAC 908

RESULT 7
 EX331615
 LOCUS
 DEFINITION EX331615 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
 cDNA clone CS0DB009Y001 5-PRIME, mRNA sequence.

ACCESSION EX331615
 VERSION EX331615.1 GI:30310073
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE 1 (bases 1 to 923)
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segr@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 7238.f. For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DB009AB01Q1&cluster=7238.f. Contact :
 Peng Liang Email: liang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Paraday Avenue Genoscope sequence ID : CS0DB009AB01Q1P1.

FEATURES

source

1..923
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /mol_type="mRNA"
 /clone="CS0DB009Y01"
 /issue_type="NEUROBLASTOMA COT 10-NORMALIZED"
 /clone_id="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.:	8,49e-126	Length:	923
Score:	1225.00	Matches:	221
Percent Similarity:	93.64%	Conservative:	0
Best Local Similarity:	93.64%	Mismatches:	15
Query Match:	69.68%	Indels:	0
		Gaps:	0

US-10-010-050A-2_COPY_31_346 (1-316) x BX331615 (1-923)

QY 1 ArgValSerGlyIleProSerArgArgHisTrpProValProCylTrpArgPheAspPhe 20
 Db 208 CGGGCTCGGGCAGTCCCTCCCGGGCCACGCGCGCGCTTCAACCGCTTGACTTC 267
 QY 21 ArgProIysProAspProTyrCysGlnAlaValTyrThrPheCysProThrGlySerPro 40
 Db 268 CGTCCAAAACCTGAGCTTATGTGACGTAAAGTATACCTTCTGCTCAACGCTGACCT 327
 QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrp 60
 Db 328 ATCCCACTTATGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 387
 QY 61 GluPheIysTyrGlyAspLeuLeuGlyHisIleuValIleMetHisAspAlaIleGlyPhe 80
 Db 388 GAATTTAAATATGAGACCTCTCTGGACACTTGAATAATATATGATGATGATGATGATG 447
 QY 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly 100
 Db 448 AGAAGTACTTAACTGGAGAACTAATACATGATGATGATGATGATGATGATGATGATG 507
 QY 101 AsnCysThrPheProHisIleuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
 Db 508 AANTGAAATTTCCCTCCCTCCGNNCTGNAATGAGTSSCCTTCTGAGTATATCAAGG 567
 QY 121 AlaAlaCysPhePheGluGlyIleAspAspValHisTrpIlyGlnAsnGlyThrLeuVal 140
 Db 568 GSTGCTGCTTTTTRGGGAATGTGTGATGTTCACTGGAAGGAAATGGACATTAGTT 627
 QY 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrTrpValIlyGlnAsp 160
 Db 628 CAAGTACCACTTATACAGAAACATGTTCAACCAATGCAAAAGGCGGTGAAGACGAGAC 687
 QY 161 AsnGluThrGlyIleTyrTyrGluThrTrpAsnValIlyAsnProGluIlyGlyAla 180

Db 688 AATGAAACGAGATTTATATGAGCATGATGATGATGATGATGATGATGATGATGATGATGATG 747

QY 181 GluThrTrpPheAspSerTyrAspCysSerIlyPheValIleArgThrPheAsnIlyLeu 200

Db 748 GAGACATGTTGATTTCTACGACATGTTCCAAATTTGTTAAGGACCTTTAAACAAGTTG 807

QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220

Db 808 GGTGAATTTGAGCGAGTTCAAGAACATAGAACCACTATCAAGATATTTCTTAC 867

QY 221 SerGlyIleProThrTyrTrpLeuGlyAsnGluThrSerValPheGlyPro 236

Db 868 ATGAGGAACCTATCTTATCTGGGAATGAAACATGCTGTTTGGGCCA 915

RESULT 8

AL571805/c 1201 bp mRNA linear EST 31-MAY-2003

LOCUS AL571805 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

DEFINITION clone CS0D1030Y01 3-PRIME, mRNA sequence.

ACCESSION AL571805 GI:31293196

VERSION AL571805.2

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE 1 (bases 1 to 1201)

JOURNAL Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 16, 2001 this sequence version replaced gi:12929467.

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: segr@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7238.f. For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0D1030CE01NP1&cluster=7238.f. Contact :

Peng Liang Email: liang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paraday Avenue Genoscope sequence ID : CS0D1030CE01NP1.

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0D1030Y01"

/issue_type="PLACENTA COT 25-NORMALIZED"

/clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.:	1.99e-124	Length:	1201
Score:	1214.00	Matches:	230
Percent Similarity:	96.64%	Conservative:	0
Best Local Similarity:	96.64%	Mismatches:	7
Query Match:	69.06%	Indels:	3
		Gaps:	0

US-10-010-050A-2_COPY_31_346 (1-316) x AL571805 (1-1201)

QY 80 PheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeu 99
 Db 1067 TTCAGAGTACTTAACTGGGCAATG-TACMAATGATGTGTATGA-MTTTCCAACCTT 1010
 QY 100 GlyAsnCysThrPheProHisIleuArgProGluMetAspAlaProPheTyrCysAsnGln 119

RESULT 10
BI090566 869 bp mRNA linear EST 20-JUN-2001
LOCUS 60285673F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996891 5',
DEFINITION mRNA sequence.
ACCESSION BI090566
VERSION BI090566
KEYWORDS EST, BI090566.1 GI:14508896
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH_MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@gsf.com, r-str@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLN at:
http://image.llnl.gov
Plate: L1AM1023 row: a column: 20
High quality sequence stop: 843.

FEATURES
source
Location/Qualifiers
1..869
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4996891"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_id="NIH_MGC_10"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo 3T.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

ORIGIN

Alignment Scores:

Pred. No.: 4.87e-122 Length: 869
Score: 1191.00 Matches: 232
Percent Similarity: 93.93% Conservative: 0
Best Local Similarity: 93.93% Mismatches: 8
Query Match: 67.75% Indels: 8
DB: 12 Gaps: 0

US-10-010-050a-2_COPY_31_346 (1-316) x BI090566 (1-869)

QY 1 ArgValSerGlyIleProSerArgArgHisTTPProValProTyrLysArgPheAspPhe 20
DB 142 CGGGTCTCGGGGATCCCTCCCGGCGCAGCTGGCGGTGCTTACAGCGCTTACATTC 201
QY 21 ArgProIysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40
DB 202 CCTCCAAACCTGATCTTATGTCAGAGTAAGTAACTTCTGTCACACTGGCTCACCT 261
QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgGluGlnAlaProValTTP 60
DB 262 ATCCAGATTTATGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 321
QY 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysGlyIleMetHisAspAlaIleGlyPhe 80
DB 322 GAATTAATATATGAGAGACCTCCGCGACACTGGAATAATATATGATGATGATGATGATG 381
QY 81 ArgSerThrLeuThrClyLysAsnTyrThrMetGluTTPTrpGlyLeuPheGlnLeuGly 100
DB 382 AGAAGTACATTAATGAGCAAGAACTACCAATGAGAAATGATGATGATGATGATGATGATG 441

QY 101 AsnGlyThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
DB 442 AACTGTACATTTCCCACTCCGACCTGAATGATGATGATGATGATGATGATGATGATGATG 501
QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTTPProTyrGluAsnGlyThrLeuVal 140
DB 502 GCTGCTCTCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 561
QY 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrVallyGlnLysP 160
DB 562 CAAGTACCAATATATGAGAGACATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 620
QY 161 AsnGluThrGlyIleTyrTyrGluThrPheAsnValLysAlaSerProGluLysGlyAla 180
DB 621 AATGAAACGGAATCTATATGAGACATGAGAAATGAGAAATGAGAAATGAGAAATGAGAA 680
QY 181 --GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysL 200
DB 681 AGAGAGATGCTTATGATCTTCTGACACTGTTCCAAATTTGTTAGAGACCTTTAACAGT 740
QY 200 euAlaGlu-PheGlyAlaGluPheLysAsnIle-GluThrAsnTyrTThrArgIlePhe 219
DB 741 TGGCTGAACTTGGGAGCAGAGTTCAGAAACATTAGAAACCAANTATACAGAAATATTTCT 800
QY 219 uTyrSerGly-GluProThrTyrLeuGly-AsnGluThr-SerValPheGlyProThrG 238
DB 801 TTACAGTGGAAAGAACTACTTATCTGGAAATGAAACCATCTGTTAGAGGGCAACCAG 860
QY 238 YAsnLys 240
DB 861 AATCAAG 867

RESULT 11
BX370641 908 bp mRNA linear EST 08-MAY-2003
LOCUS BX370641 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL004YP03 5-PRIME, mRNA sequence.
ACCESSION BX370641.1 GI:30459737
VERSION EST.
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 908)
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7338.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG053ZH03_CS05016_1&cluster=7238.f.
Contact : Feng Liang Email : fliang@life.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG053ZH03_CS05016_1.

FEATURES
source
Location/Qualifiers
1..908
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL004YP03"
/cell_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and Ecor V

1. 908
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL004YP03"
/cell_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and Ecor V

sites of the PCMVSPORT 6 vector. Library was normalized.

Alignment Scores:	
Prod. No.:	2,44e-119
Matches:	167.00
Conservative:	96.88
Mismatches:	96.88
Indels:	65.38
Gaps:	13

JS-10-010-050A-2_COPY_31_346 (1-316) X BX370641 (1-908)

2Y	93	TrpTyrGluLeuPheGlnLeuGlyAsnGlyThrPheProHisLeuLysProGluMetAsp	1..2
2b	4	TGGTCTGGACATTTTTCACCTTGGGCAACTGGTACATTTCCCATCTCCGACTGA-ATGGAT	6..2
2Y	113	AlaProPheThrCysAsnGlnGlyAlaAlaCysPhePheGluGlyIleAspAspValHis	1..32
2b	63	GCCCCCTTCTGGGTGATATCAAGCGCGTCCCTGCTTTTGGAGGAATGATGATGTTCAAC	1..32
2Y	133	TrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln	1..32
2b	123	TGGAGGAAATATGGACATTAGTTCAAGTGCACATATACAGGAAACATGTTCAACAA	1..32
2Y	153	MetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyrTrpGluThrTrpAsnVal	1..72
2b	183	ATGGCAAAAGTGGGTGAAACAGGACAAAGGAAATTTATTTATGAGACATGGAAATGA	2..42
QY	173	LysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTrpAspCysSerLysPhe	1..92
Db	243	AAACCCGACCCAGAAAGGGGGCAGAGCATGGTTGATTTCTTCGACGTGTCCAAATTT	3..02
QY	193	ValLeuArgTrpPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThr	2..12
Db	303	GTCGTTAAGACCTTTAACAACTGGCTGGAATTTGAGACAGAGTTCAGAAACATGGAACC	3..62
QY	213	AsnTyrThrTrpArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSer	2..32
Db	363	AACTATACAAAGAAATATTTCTTTTAAAGTGGAAACCTATCTTTCGGGAAAGAAACATCT	4..22
QY	233	ValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTrpPro	2..52
Db	423	GTTTTTGGGCAACAGGAAACAGACTTGTGGTTTATGCCATTAATAAAGATTTTATACCC	4..82
QY	253	PheLysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAla	2..72
Db	483	TTCAATCCACATTTGGCCAACTTAAGAAATTTCTGTGAATCTCTTGCAGAAATTTGATGCA	5..42
QY	273	ValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeuProMet	2..92
Db	543	GTCATTTGGCAACAAACAGTTCTATTTGTTTATATATTTGGAATATGTGGTTTATACCTATG	6..02
QY	293	LysPheProPheIleLysIleThrTyrGluGluIleProLeuProLysArgAsnLysThr	3..12
Db	603	AAATTCCTCTTATTAATAATACATATGAGAAATCCCTTATCCTATCAGAAACAAACAA	6..62
QY	313	LeuSerGlyLeu 316	
Db	663	CTCTCTGGTTTA 674	

RESULT 12			
B0919385			
LOCUS	B0919385	974 bp	mRNA
DEFINITION	AGENCOCRT_8802804 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6382620		linear EST 20-AUG-2007
	5', mRNA sequence.		

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 974)	NIH-MGC	http://mgc.ncbi.nlm.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.			
	Email: strausbe-r@nhih.gov			

FEATURES

SOURCE

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6382620"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site.1: XhoI;
Site.2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

```

ORIGIN

US-10-010-010-050A-2 COPY 31 346 (1-316) X BQ919385 (1-974)

Dy 7 SetArgArgHisIstIrrProValProTyrIlysArgPheAspPheArgProIysProAspPro 26
:::
117 TCGGGCAACGGCTGGCGGGTGCCCTTCAAGGGCTTCTTTCCGTCCGAACACATATCC 176

Dy 27 TyrGlnAlaIalysTyrThrPheCysProIrrGlySerProIleProValMetGluGly 46
:::
177 TACTGCAAGCTAAGATACCTTCTGTCTTACCAGGCTGCGCCATCCCACTTATGAAGAC 236

Dy 47 AspAspAspIleGluValPheArgLeuGlnAlaProValTrrGluPheIysTrrGlyAsp 66
:::
237 AATGACGTACATGAGGCTCTTACGACTACCAAGCCCGCATTTGGAAATTTAAATATGAGAC 296

Dy 67 LeuIleGluYHileuIysIleMetHisAspAlalleGlyPheArgSerThrIrruThrGly 86
:::
297 CTCCTGGGACACTTAACTTATGACATGACCGCTGGGATTCAGAGGACACTGACAGGC 356

Dy 87 LysAsnTyrThrMetGluTrrGlyGluLeuPheGlnLeuGlyAsnGlyThrPheProHis 106
:::
357 AAGATCTACACATTAGGTGATGAACCTTTCCACTGGGCAACGTACATTTCCCCAC 416

Dy 107 LeuArgProGluMetAspAlaProPheTrrPyrAsnGlnGlyAlaIalCysPhePheGlu 126
:::
417 CTCGGGCTACACAAGCGCTCCCTCTGTGTTGAACCAAGGGCAGCTGCTTTTGGAA 476

Dy 127 GlYIleAspAspValHisTrrPylsGluAsnGlyThrLeuValGlnValAlaThrIleSer 146
477 GGAATATGATATTAACACTGGAGGAAACCGGACACTGTCAGTGGTGCACCATATCC 536

Dy 147 GlyAsnMetPheAsnGlnMetAlaIalysTrrValIysGlnAspAsnGluThrrGlyIleTyr 166

```

Db      537  GGAAGCATTATACAAAGTGGCCAGTGGTGAACAGAGCAATGAACTGGGATTAT 596
Qy      167  TTTGTTTCTTTPAenVallysalaserProGluLysGlyAlaGluThrTPheaspSer 186
Db      597  TATGAGACATGAGACGGTCCAGCCGAGCAGCAAGGGGCCAGAGCGTGTTCGAGTCC 656
Qy      187  TTTAspCysSerlyspheValleuArgThrPheasnlyseuAlaGluPheGlyAlaGlu 206
Db      657  TACGACTGTTGCAATTTTGTCTTAAAGACATATAGAAATTTGGCTGATTTGGAACAGAA 716
Qy      207  PheylsasnilleGluThrAsnlyrThrArgilepheleuTyserGlyGluProThrTyr 226
Db      717  TTCAAGAAATGAGAAACAACTATACGAAATATTTCTTACAGTGAAGAGCTATTTAC 776
Qy      227  LeuGlyasnGluThrSerValPheGlyProThrGlyAlasnlyserThr-leuGlyleuAlaI 246
Db      777  CTGGGAATGAAACATCTATTTTGGGCCCAAGAGGAAAGACTTGTGCTTGGCCAT 836
Qy      246  e-LysAspPheTyTYrTyProPheTyProHisleu---ProThrTyserGlu-Pheleu 264
Db      837  AAAAAAATTTTATGTCCTCCCTCCAGACCGAATTTGTCAACCAAGATTTTCCTGATG 896
Qy      265  SerleuLeuGln---llePheaspAlaVal-----lleVal 275
Db      897  AAATTTCTTGAATAATTTTGTATACGTGATATATACCCCGAGGTTCTTACTTGT 956
Qy      276  HisTyserGlnPheTyR 280
Db      957  TTTAAACCTTTGG 971

RESULT 13
LOCUS      BI080303          915 bp      mRNA      linear      EST 20-JUN-2001
DEFINITION 602876736f1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5008544 5',
            mRNA sequence.
VERSION     BI080303
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
REFERENCE   1  (bases 1 to 915)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/MLNL at:
            http://image.jnl.gov
            Plate: L14M1053 row: 9 column: 09
            High quality sequence stop: 748.
            Location/Qualifiers
              1..915
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="FVB/N-3"
                /db_xref="taxon:10090"
                /clone="IMAGE:5008544"
                /tissue_type="tumor, biopsy sample"
                /dev_stage="5 months"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigator:
                providing samples: Gilbert Smith, NIH"
FEATURES
  source

```

```

ORIGIN
Alignment Scores:
Pred. No.:      2,48e-118
Score:          1158.00
Percent Similarity: 84.83%
Best Local Similarity: 75.86%
Query Match:    65.87%
DB:             12
               Gaps: 1

US-10-010-050A-2_copy_31_346 (1-316) x BI080303 (1-915)
Qy      25  AspproTyTyCyGlnAlaValTyThrPheCyProThrGlySerProIleProValMet 44
Db      25  GATCCTTACTGTCAGACTAGTACTTCTTCTTACCGGCTCGCCATCCAGTTATG 84
Qy      45  GluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTTPGluPheTyR 64
Db      85  AAGAGACATGAGCTCATCGAGGTCTTACGACTACAGCCCGCATTTGGGAATTAATAT 144
Qy      65  GlyAspLeuLeuGlyHisleuYsilleMetHisAspAlalleGlyPheArgSerThrLeu 84
Db      145  GGAGACCTCTCGGACACTTAACTTATGATGACAGCGCTGGGATTCAGAGCACTG 204
Qy      85  ThrGlyTyAsnTyThrMetGluTrpTyGluLeuPheGlnLeuGlyAsnTyThrPhe 104
Db      205  ACGAGCAAGAACTACACAAATGAGTGTATGATCTTTCCAGCTGGGCAACTGATATT 264
Qy      105  ProHisleuArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhe 124
Db      265  CCCACCTCGGCTGAGCAAGAGCGCTCCTTGTGGTGTAAACCAAGGGGAGCTGCTTT 324
Qy      125  PheGluGlyTlleAspAspValHisTrpTyserGluAsnGlyThrleuValGluValAlaThr 144
Db      325  TTTGAAGATATAGATGATTAACCTGAGAAAGGAAAGCGGACCTGTCAGTCTTGCAAC 384
Qy      145  IleSerGlyAsnMetPheAsnGlnMetAlaTyserTyVallyserGlnAspAsnGluThrGly 164
Db      385  ATATCCGAAACACATTTAAACAAAGTGGCCGAGTGGTGAAGAGCAATGAACTGGGG 444
Qy      165  IleTyTyGluThrTPAenVallysalaserProGluTyserGlyAlaGluThrTPhe 184
Db      445  ATTATTTATAGACATGACGAGTCCGAGCGGCCAGCAAGGGGCCAGACCTGTGTC 504
Qy      185  AspSerTyAspCysSerlyspheValleuArgThrPheasnlyseuAlaGluPheGly 204
Db      505  GAGTCTTACGACTGTTGAAATTTGTCTTAAAGACATATAGAAATCGCTGAATTTGGA 564
Qy      205  AlaGluPheLysAsnilleGluThrAsnTyThrArgilepheleuTyserGlyGluPro 224
Db      565  ACGAATTCAGAAAGATGAAACAACTATACGAAATTTCTTACAGTGAAGAGCT 624
Qy      225  ThrTyrlleuGlyAsnGluThrSerValPheGlyProThrGlyAlasnlyserThrleuGly 244
Db      625  ATTATACCTGGGAATGAAACATCTATTTTGGGCCCAAGGAAACAGACTGTGCTTGG 684
Qy      245  AlaIlelybArgPheTyTYrTyProPheTyProHisleuProThrTyserGlyPheleu 264
Db      685  GCCATAAAAAACTTATAGCCCTTCAGACCGTA-TTGTCAACCAAGATGTTCTGATG 743
Qy      265  SerleuLeuGlnillePheaspAlaValleValHis-LysGlnPheTyrlleuPheTyRas 284
Db      744  AATTC-TTGACAACTTTGATACAGTATTTTACCGCAAGAGCTACTGCTTTATATA 802
Qy      284  nPheGluTyTyThrPheleuProMetLysPheProPheIleTyserIleThrTyTyGluGlu 304
Db      803  CTTGAGATA-TGGGTTCACCATGAAA---CCCTTGTGCAATTAATACATGAGAGAAAC 858
Qy      304  eProleuProIleArgAsnlyserThrleu 313
Db      859  -CCGGTTACGTACGAAATTCAGGTTA 865

RESULT 14

```

```

distribution: NCI-CGAP clone distribution information can be
through the I.M.A.G.E. Consortium/ILN at:
     1. gov
     2. CAM3163 row: E column: 14
     3. M3RPI reverse primer (AB1).
     4. location/Qualifiers
     5. 1..641
     6. /organism="Homo sapiens"
     7. /mol_type="mRNA"
     8. /db_xref="taxon:9606"
     9. /clone="IMAGE:5937013"
    10. /sex="female"
    11. /tissue_type="normal endometrium, late proliferative
        phase, cycle day 13"
    12. /lab_host="DH10B (TI-resistant)"
    13. /clone_id="NCCHD_HS_Ut1"
    14. /name="Organ: uterus; Vector: pCMV-SPORT6.1.cdb (Resgen,
        Invitrogen Corporation); Site 1: NotI; Site 2: EcoRV;
        Cloned unidirectionally from microquantity amounts of mRNA
        from normal endometrial tissue (late proliferative phase,
        cycle day 13). Average insert size 1.9 kb. Library
        constructed by Resgen (Invitrogen Corporation)."
```

Alignment Scores:

Pred. No.:	4.52e-116	Length:	641
Score:	1136.00	Matches:	200
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	64.62%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050a-2_copy_31_346 (1-316) x CB215456 (1-641)

```

QY      1  ArgValSerGlyIleProSerArgArgHisTrpProValProTyrIleArgPheAspPhe 20
          |||
Db      42  CGGCTCGGGGCATCCCTCCCGGGCCACCTGGCCGGCTTCAAGCGCTTGACTTC 101
          |||

QY      21  ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40
          |||
Db     102  CGTCCAAACCTGATCTTATTTGACAGCTAGATACTTCTGTCTCAACTGGCTCACT 161
          |||

QY      41  IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrp 50
          |||
Db     162  ATCCCAAGTTATGGAGGGTGAATGATGACATTAAGTTTTCATTACAAGCCCGATATGG 221
          |||

QY      61  GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe 90
          |||
Db     222  GAATTTAAATGTGAGACCTCTCTGGACACTTGAAATTTATGCATGATGCCATTGGATTTC 281
          |||

QY      81  ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly 100
          |||
Db     282  AGAAGTACATTAATCTGGCAGACACTACACAAATGGAATGGATGTAACCTTTCCAACTTGGC 341
          |||

QY     101  AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
          |||
Db     342  AACTGTACATTTCCCATCTCCGACCTGAATGATGATGCCCTTCTGTGTATCAAGGC 401
          |||

QY     121  AlaAlaCysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuVal 140
          |||
Db     402  GCTGCCCTGCTTTTGGAGGGAATGATGATGTCTCACTGGAAGGAAATGGGACATTAGTT 461
          |||

QY     141  GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsp 160
          |||
Db     462  CAAAGTACACTATATACGAAACATGTTTCAACCAAAATGGCAAAATGGGTGAACAGGAC 521
          |||

QY     161  AsnGluThrGlyIleTyrTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180
          |||
Db     522  AATGAAACAGGAATTTATATGAGACATGGAATGTAAAGCCAGCCCAAAAAGGGGCA 581
          |||

QY     181  GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200
          |||
Db     582  GAGCAATGGTTTGAATCTTCAAGACTGTCCAAATTTGTATAGGACCTTTAAACAAAGTTG 641
          |||

```

Search completed: April 25, 2004, 05:46:13
 Job time : 4028.01 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 23:04:07 ; Search time 629.806 Seconds
(without alignments)
2262.168 Million cell updates/sec

Title: US-10-010-050A-2_COPY_31_346

Perfect score: 1/58
Sequence: 1 RVSGIPSRHMPVVPYKRFDF.....IKITYEETPLDINKKTSGL 316

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US10010050/runat_22042004_113205_27584/app_query.fasta_1.1372
-DB=published Applications NA -QFMT=fastap -SUFFIX=rmpb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -STRAT=1 -END=1 -MATRIX=blosum62
-TRANS=human40.ccd -LIST=45 -DOCLALIGN=200 -THR SCORE=pcp -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcp -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USPR=US10010050 @CEN 1 1 776 @runat_22042004_113205_27584
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSEBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1758	100.0	1486	9	US-09-122-383-1
2	1758	100.0	1486	14	US-10-010-050A-1
3	1747	99.4	1751	13	US-10-653-595-26
4	1747	99.4	1751	13	US-09-397-945-26
5	1515	86.2	1038	9	US-09-122-383-13
6	1515	86.2	1038	14	US-10-010-050A-13
7	886	50.4	494	9	US-09-864-761-22592
8	727	38.9	506	14	US-10-040-739-825
9	684.5	31.9	474	9	US-09-864-761-5832
10	472	26.8	497	16	US-10-264-049-1859
11	359.5	20.4	351	9	US-09-867-701-2579
12	241	13.7	473	9	US-09-954-456-1363
13	121	6.9	60	10	US-09-908-975-10339
14	104.5	5.9	2667	15	US-10-198-846-9753
15	97.5	5.5	986	16	US-10-359-493-36401
16	96	5.5	3580	10	US-09-828-846-21
17	96	5.5	3580	10	US-09-374-046A-159
18	96	5.5	3580	13	US-10-616-263-159
19	94.5	5.4	14070	9	US-09-870-759-127
20	94.5	5.4	14070	10	US-09-751-708A-127
21	94	5.3	3717	13	US-10-282-122A-22565
22	92.5	5.3	1612	14	US-10-003-392-9
23	92.5	5.3	2115	14	US-10-003-392-19
24	92	5.2	1468	13	US-10-424-599-120013
25	91.5	5.2	7680	10	US-09-953-318-3
26	91.5	5.2	7680	13	US-10-211-462-214
27	91.5	5.2	7680	13	US-10-342-887-711
28	91.5	5.2	7680	13	US-10-172-118-711
29	91.5	5.2	7680	13	US-10-446-373-9
30	91.5	5.2	7680	15	US-10-021-660-16
31	91.5	5.2	7680	15	US-10-393-590-77
32	91.5	5.2	7680	15	US-10-393-567-77
33	91.5	5.2	7680	15	US-10-394-087-77
34	89	5.1	2131	13	US-10-424-599-109423
35	88.5	5.0	870	16	US-10-369-493-26859
36	88.5	5.0	3035	16	US-10-108-260A-1791
37	88.5	5.0	5574	15	US-10-037-270-253
38	88.5	5.0	5574	16	US-10-117-723-253
39	88	5.0	1317	15	US-10-028-245-4
40	88	5.0	1815	13	US-10-282-122A-23914
41	88	5.0	1826	15	US-10-028-245-1
42	88	5.0	9585	13	US-10-282-122A-22862
43	87	4.9	1167	16	US-10-369-493-43789
44	87	4.9	1314	13	US-10-282-122A-16281
45	87	4.9	3488	16	US-10-365-493-36547

ALIGNMENTS

RESULT 1
US-09-122-383-1
Sequence 1, Application US/09122383A
Patent No. US20020042093A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38
CURRENT APPLICATION NUMBER: US/09/122,383A
CURRENT FILING DATE: 1998-07-24
EARLIER APPLICATION NUMBER: 60/053,613
EARLIER FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1486
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (47)...(1084)

Sequence 1, Appl
Sequence 1, Appl
Sequence 26, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 825, A
Sequence 5832, A
Sequence 1859, A
Sequence 2579, A
Sequence 1363, A
Sequence 10339, A
Sequence 36401, A
Sequence 21, Appl
Sequence 159, A
Sequence 127, A
Sequence 127, A
Sequence 2265, A
Sequence 9, Appl
Sequence 19, Appl
Sequence 120013, A
Sequence 3, Appl
Sequence 214, A
Sequence 711, A
Sequence 711, A
Sequence 16, Appl
Sequence 77, Appl
Sequence 77, Appl
Sequence 77, Appl
Sequence 109423, A
Sequence 26859, A
Sequence 1791, A
Sequence 253, A
Sequence 253, A
Sequence 4, Appl
Sequence 23914, A
Sequence 1, Appl
Sequence 22862, A
Sequence 43789, A
Sequence 16281, A
Sequence 36547, A

US-09-122-383-1

Alignment Scores:

Pred. No.:	2,286-222	Length:	1486
Score:	1758.00	Matches:	316
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-010-050A-2_COPY_31_346 (1-316) x US-09-122-383-1 (1-1486)

```

QY 1 ArgValSerGlyLleProSerArgAghIstPProValProTyLysArgPheasphe 20
Db 137 CGGGTTCGGGCAATCCCTCCCGGCCCACTGGCCGGTCCCTCAAGAGCCTTGACTTC 196
QY 21 ArgProLysProAspProTyTyCysGlnAlaLysTyThrPheCysProThrgLysPro 40
Db 197 CGTCCAAAACCTGATCTTATGTCAAGCTAAGTATACCTTCTGTCCAACTGGCTCACCCT 256
QY 41 IleProValMetGlnGlyAspAspApIleGluValPheArgGluGlnAlaProValTTrp 60
Db 257 ATCCCAAGTATGAGGGGTGATGATGACATTGAAGTTTTCGATTACAGCCCAAGTATG 316
QY 61 GluPheLysTyrgLysAspLeuGlyHsLeuLysIlleMetHsAspAlaIleGlyPhe 60
Db 317 GAATTTAAATATGAGACCTCTCGGACACTTGAAATATATGATGATGATGATGATG 376
QY 81 ArgSerThrLeuThrgLysAsnTyThrMetGluTrpTyGluLeuPheGlnLeuGly 100
Db 377 AGAAGTATACATTAATCTGGCAAGAACTACACATGGAATGGATGATGATGATGATG 436
QY 101 AsnCySerThrPheProHsLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
Db 437 AACGTACATTTCCCGCATCTCCGACCTGGAATGATGATGATGATGATGATGATG 496
QY 121 AlaAlaCysPhePheGlnGlyLleAspAspValHsTrpLysGluGlnGlyThrLeuVal 140
Db 497 GCTGCCCTGCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATG 556
QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsp 160
Db 557 CAATAGCAACTATATACGAAACATGTTCACCAATAGCAAGTGGCAAGTGGCAAGTGG 616
QY 161 AsnGluThrgLylleTyTyTyGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180
Db 617 AATGAAACAGGAATTTATATGAGACATGATGATGATGATGATGATGATGATGATG 676
QY 181 GluThrTrpPheAspSerTyAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200
Db 677 GAGACATGCTTGTGATCTCTAGACTGTTCCAAATTTGGTTAAGACCTTTAACAAGTGG 736
QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyThrAspGlnPheLeuTy 220
Db 737 GCTGAATTTGAGCAGAGTTCAGAACATAGAACCAACTATACAAAGATTTTCTTTAC 796
QY 221 SerGlyLysProThrTyLeuGlyAsnGluThsSerValPheGlyProThrGlyAsnLys 240
Db 797 AGTGAAGACCTACTTACTCTGAGAAATGAAACATCTGTTTGGCCCAACGAGAAACAG 856
QY 241 ThrLeuGlyLysAlaIleLysArgPheTyTyTyProPheLysProHsLeuProThrls 260
Db 857 ACTCTGCTTGGTTCAGCAATGAAATGATTTATACCCCTTCAACCAACATTTCCAACTAAA 916
QY 261 GluPheLeuSerLeuLeuGlnIlePheAspAlaValIleValHsLysGlnPheTy 280
Db 917 GAATTTCTGTTGAGTCTCTGCAAAATTTTGTGATGAGATGATGATGATGATGATGATG 976
QY 281 LeuPheTyTrpAsnPheGluTyThrPheLeuProMetLysPheProPheIleLysIleThr 300
Db 977 TTGTTTAAATTTGAAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1036
QY 301 TyrgLysLysIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 316

```

Db 1037 TATGAGAAATCCCTTATCTATCATGAAACAAACACTCTCTGCTTGA 1084

RESULT 2

US-10-010-050A-1

```

; Sequence 1, Application US/10010050A
; Publication No. US20020173624A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; TITLE OF INVENTION: CHROMOSOME 13
; FILE REFERENCE: 97-38C1
; CURRENT APPLICATION NUMBER: US/10/010,050A
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 09/122,383
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/053,613
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (47)...(1084)
US-10-010-050A-1

```

Alignment Scores:

Pred. No.:	2,286-222	Length:	1486
Score:	1758.00	Matches:	316
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050A-2_COPY_31_346 (1-316) x US-10-010-050A-1 (1-1486)

```

QY 1 ArgValSerGlyLleProSerArgAghIstPProValProTyLysArgPheasphe 20
Db 137 CGGGTTCGGGCAATCCCTCCCGGCCCACTGGCCGGTCCCTCAAGAGCCTTGACTTC 196
QY 21 ArgProLysProAspProTyTyCysGlnAlaLysTyThrPheCysProThrgLysPro 40
Db 197 CGTCCAAAACCTGATCTTATGTCAAGCTAAGTATACCTTCTGTCCAACTGGCTCACCCT 256
QY 41 IleProValMetGlnGlyAspAspApIleGluValPheArgGluGlnAlaProValTTrp 60
Db 257 ATCCCAAGTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 316
QY 61 GluPheLysTyrgLysAspLeuGlyHsLeuLysIlleMetHsAspAlaIleGlyPhe 80
Db 317 GAATTTAAATATGAGACCTCTCGGACACTTGAAATATATGATGATGATGATGATG 376
QY 81 ArgSerThrLeuThrgLysAsnTyThrMetGluTrpTyGluLeuPheGlnLeuGly 100
Db 377 AGAAGTATACATTAATCTGGCAAGAACTACACATGGAATGGATGATGATGATGATG 436
QY 101 AsnCySerThrPheProHsLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
Db 437 AACGTACATTTCCCGCATCTCCGACCTGGAATGATGATGATGATGATGATGATG 496
QY 121 AlaAlaCysPhePheGlnGlyLleAspAspValHsTrpLysGluGlnGlyThrLeuVal 140
Db 497 GCTGCCCTGCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATG 556
QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsp 160
Db 557 CAATAGCAACTATATACGAAACATGTTCACCAATAGCAAGTGGCAAGTGGCAAGTGG 616
QY 161 AsnGluThrgLylleTyTyTyGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180

```

|||||
617 AATGAAACAGAAATTTATATGACACATGAAATGTAAGCCACCCAGAAAAGGGGGCA 676
|||||
181 GlnThrTrpPheAspSerTyrAspCysSerIysPheValIleuArgThrPheAsnIleu 260
|||||
677 GAGCAATGGTTGATTCCTACGACTGTTCCAAATTTGTGTTAGACCTTTAAACAATTG 726
|||||
201 AlaGluPheGlyAlaGluPheIysAsnIleGluThrAsnIleThrArgIlePheLeuTyr 220
|||||
737 GCTGAATTTGGAGCAGATTCAAGACATGAAACCAACATACAAAGAATATTTCTTTAC 756
|||||
221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIys 240
|||||
797 AGTGAGAACTACTTATCTCGGAAATGAAACATCTGTTTGGCCCAACGGAACAG 856
|||||
241 ThrIleuGlyLeuAlaIleIysArgPheTyrTyrProPheIysProHisIleProThrIys 260
|||||
857 ACTCTGGTTTACGCCAATAAAAGATTTTATTAACCCCTTCAAAACACATTTGCCAATCAA 926
|||||
261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisIysGlnPheTyr 280
|||||
917 GAATTTCTGTTGAGCTCTTGCAAAATTTTGATGACGATGATGCAACAACAGTTCTAT 976
|||||
281 LeuPheTyrAsnPheGluTyrTyrPheLeuProMetIysPheProPheIleIysGleThr 300
|||||
977 TTGTTTATTAATTTTGAATATGTTTGTATTAACCTATGAAATTCCTTTATTAATAATAACA 1036
|||||
301 TyrGluGluIleProLeuProIleArgAsnIysThrLeuSerGlyLeu 316
|||||
1037 TAGGAAGAAATCCCTTACCTTACCTACGAAACAAACACATCTCTGTTTA 1084

RESULT 3

US-10-653-595-26

; Sequence 26, Application US/10653595

; Publication No. US20040048304A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et. al.

; TITLE OF INVENTION: 95 Human secreted proteins

; FILE REFERENCE: P2027PICI

; CURRENT APPLICATION NUMBER: US/10/653,595

; CURRENT FILING DATE: 2003-09-03

; PRIOR APPLICATION NUMBER: US 09/397945

; PRIOR FILING DATE: 1999-09-17

; PRIOR APPLICATION NUMBER: PCT/US99/05804

; PRIOR FILING DATE: 1999-03-18

; PRIOR APPLICATION NUMBER: 60/078,566

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/078,576

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/078,573

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/078,574

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/078,579

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/080,314

; PRIOR FILING DATE: 1998-04-01

; PRIOR APPLICATION NUMBER: 60/080,312

; PRIOR FILING DATE: 1998-04-01

; PRIOR APPLICATION NUMBER: 60/078,578

; PRIOR FILING DATE: 1998-03-19

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 470

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 26

; LENGTH: 1751

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (1520)

; OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-10-653-595-26

Alignment Scores:

Pred. No.: 8,57e-221 Length: 1751
Score: 1747.00 Matches: 314
Percent Similarity: 99.37% Conservative: 0
Best Local Similarity: 99.37% Mismatches: 2
Query Match: 99.37% Indels: 0
DB: Gaps: 0

US-10-010-050A-2_COPY_31_346 (1-316) x US-10-653-595-26 (1-1751)

QY 1 ArgValSerGlyIleProSerArgArgHisThrProValProTyrIleArgPheAspPhe 20
DB 139 CGGGTCTGGGCATCCCTCCCGGGCCACGCGCGGCTTACAGCGCTTTACATTC 198
QY 21 ArgProIysProAspProTyrCysGlnAlaIysTyrThrPheCysProThrIleYserPro 40
DB 199 GCTCCAAACCTGATCCTTATTTGCAAGTATATCTTCTGTCGCAACTGGCTCACCT 258
QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValIlePro 60
DB 259 ATCCAGTTATGAGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 318
QY 61 GluPheIleTyrGlyAspLeuLeuGlyHisLeuValIleMetHisAspAlaIleGlyPhe 80
DB 319 GAATTTAATATGAGACCTCTCGGACACTTGAACATTAATGACATGATGATGATGATG 378
QY 81 ArgSerThrLeuThrGlyIleAsnIleThrMetGluTyrTyrGluLeuPheGlnLeuGly 100
DB 379 AGAAGTACATTACCTGCGAAGAACTACCAATGATGATGATGATGATGATGATGATGATG 438
QY 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheThrCysAsnGlnGly 120
DB 439 AACTGTACATTTCCCATCTCCGACCTGAAAGATGCCCTTCTGCTGATGATGATGATG 498
QY 121 AlaIleCysPhePheGluGlyIleAspAspValHisThrIleGluValIleLeuVal 140
DB 499 GCTGCTGCTTTTGTGAGGAATTTGATGATGATGATGATGATGATGATGATGATGATG 558
QY 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleSerTyrValIysGlnAsp 160
DB 559 CAAAGTACCACTATACGAAACATGTTCAACCAATGCAAAAGGCGGTGAACAGGAC 618
QY 161 AsnGluThrGlyIleTyrTyrGluThrTrpAsnValIleAlaSerProGluIysGlyAla 180
DB 619 AATGAACACAGAAATTTATATGACATGAAATGTAAGCCACCCCAAAAAGGGGGCA 678
QY 181 GluThrTrpPheAspSerTyrAspCysSerIysPheValIleuArgThrPheAsnIleu 200
DB 679 GAGACATGGTTGATTCCTACGACTGTTCCAAATTTGTGTTAGACCTTTAAACAATTG 738
QY 201 AlaGluPheGlyAlaGluPheIysAsnIleGluThrAsnIleThrArgIlePheLeuTyr 220

Db 739 GGTGAATTGGAGCAGAGTTCAAGAACATGAAACCACTATACAGATATTTCTTAC 798
Qy 221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
Db 799 AGTGGAGAACCTACTATCTGGAATGAAACATCGTTTTGGGCGCAACAGAAACAG 858
Qy 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLys 260
Db 859 ACTCTTGTTTACCCCTAATAAGATTTTATACCCCTTAAACACATTTGGCAACTAA 918
Qy 261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyr 280
Db 919 GAATTCCTGTGAGTCTCTTGCATTTTGAATGACAGTATGTGCACAAACAGTTCTAT 978
Qy 281 LeuPheTyrAsnPheGluTyrTyrPheLeuProMetLysPheProPheIleLysIleThr 300
Db 979 TTGTTTATATATTTTGAATATGTTTATACCTATTAATCCCTTTTATTAATAATACA 1038
Qy 301 TyrGluGluIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 316
Db 1039 TATGAAGAAATCCCTTACTATCAGAAACAAACACTCTTGTTT 1086

RESULT 4

US-09-397-945-26
; Sequence 26, Application US/09397945
; Publication No. US20030065139A1

; GENERAL INFORMATION:

; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: 95 Human secreted proteins

; FILE REFERENCE: P2027P1

; CURRENT APPLICATION NUMBER: US/09/397,945

; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PCT/US99/05804

; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/078,566

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,576

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,573

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,574

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,579

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,314

; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080,312

; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/078,578

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,581

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,577

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,563

; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,313

; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080,313

; NUMBER OF SEQ ID NOS: 470
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 26

LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: SITE
LOCATION: (1520)

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

LOCATION: (1557)

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

LOCATION: (1689)

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-09-397-945-26

Alignment Scores:

Pred. No.:	8,57e-221	Length:	1751
Score:	1747.00	Matches:	314
Percent Similarity:	99.37%	Conservative:	0
Best Local Similarity:	99.37%	Mismatches:	2
Query Match:	99.37%	Indels:	0
DB:	13	Gaps:	0

US-10-010-050a-2_copy_31_346 (1-316) x US-09-397-945-26 (1-1751)

Qy 1 ArgValSerGlyIleProSerArgArgHisTyrProValProTyrLysArgPheAspPhe 20
Db 139 CGGGTCTGGGCAATCCCTCCCGCGCACTGGCGGCTTCAAGCGCTTGAATTC 198
Qy 21 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40
Db 199 CGTCAAAACCGAACCTTATGTCAAGTAAATGTAATCTTGTCCAACTGCTCACCCT 258
Qy 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyr 60
Db 259 ATCCCAAGTATGAGAGGTATGATGACATTAAGATTTTGCATTAACGCCCAAGATG 318
Qy 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe 80
Db 319 GAATTTAATATGAGAACCTCTCGGACACTTGAAATATATGATATGATGATGATTC 378
Qy 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGly 100
Db 379 AGAAGTACATTAATCTGCAAGAACTACACAAAGATGATGAACTTTTCCAACTTGGC 438
Qy 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
Db 439 AACTGTACATTTCCCATCTCCGACCTGAAAGGATGCCCTTTCTGGGTATATCAAGGC 498
Qy 121 AlaAlaCysPhePheGluGlyIleAspAspValHisTyrLysGluAsnGlyThrLeuVal 140
Db 499 GCTGCTGCTTTTATAGGAATGATGATGATTCATGGAAGGAAATGGGACATTAGTT 558
Qy 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrValLysGlnAsp 160
Db 559 CAAGTAGCACTATATCAGGAAACATGTTCAACCAATGGCAAAAGGGTGAACAGGAC 618
Qy 161 AsnGluThrGlyLysTyrTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180
Db 619 AATGAAACAGGAATTTATATGAGACATGGAATGTAACCCAGCCCAAGGAGGAGCA 678
Qy 181 GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200
Db 679 GAGACATGTTTATCTTCCATCGACTGTCCAAAATTTGTATAGGACCTTTAACAAGTTG 738
Qy 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220
Db 739 GCTGAATTTGAGAGAGATTCAGAAACATGAAACCAACTATACAGAAATATTTCTTAC 798
Qy 221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
Db 799 AGTGGAGAACCTACTATCTGGAATGAAACATCGTTTTGGGCGCAACAGAAACAG 858
Qy 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLys 260
Db 859 ACTCTTGTTTACCCCTAATAAGATTTTATACCCCTTAAACACATTTGGCAACTAA 918

QY 261 GIUPELEULEUSERLEULEGINTLEPHEAPALAVALLILEVALHISLYSGINPHELYR 280
 DB 919 GAATTCCTGTTGAGCTCTTGCAAAATTTTGTATGACAGGATTTGCACAAACAGTTTCAT 978
 QY 281 LEUPHELYRASPHEGLUPLYRTTRPHELEUPROMETLYSPHEPROPHETILELYSILETHR 300
 DB 979 TTGTTTAAATTAATTTGAATTAATGGTTTTCCTATGAAATTCCTTTTATTAATAATACA 1038
 QY 301 TYRGLUGINTLEPROLEUPROILEARGASNLVSTHLEUSERGLYLEU 316
 DB 1038 TATGAAGAAATCCCTTTACCTATCAGAAACAAACACCTCTGTTTA 1086
 RESULT 5
 US-09-122-383-13
 ; Sequence 13, Application US/09122383A
 ; Patent No. US20020042093A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; APPLICANT: Gilbertson, Debra G.
 ; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
 ; FILE REFERENCE: 97-38
 ; CURRENT APPLICATION NUMBER: US/09/122,383A
 ; EARLIER FILING DATE: 1998-07-24
 ; EARLIER APPLICATION NUMBER: 60/053,613
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 1038
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Degenerate nucleotide sequence encoding zs1946
 ; OTHER INFORMATION: polypeptide of SEQ ID NO:2
 ; NAME/KEY: variation
 ; LOCATION: (1)...(1038)
 ; OTHER INFORMATION: n is any nucleotide
 ; US-09-122-383-13
 Alignment Scores:
 Pred. No.: 2,276-190 Length: 1038
 Score: 1515.00 Matches: 267
 Percent Similarity: 85.03% Conservative: 47
 Best Local Similarity: 85.03% Mismatches: 47
 Query Match: 86.18% Indels: 0
 DB: 9 Gaps: 0
 US-10-010-050A-2_COPY_31_346 (1-316) x US-09-122-383-13 (1-1038)
 QY 2 ValSerGlyIleProSerArgHisTrpProValProTyrIysArgPheAspPheArg 21
 DB 94 GTTMSNGNATHTCCNWMNMGNCAYTGCCGCTGTCNCAATTAABAGNTTGGATTTTGMN 53
 QY 22 ProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerProIle 41
 DB 154 CQNAARCCNGAYCCNTATYGCARGCNAARTAYACNTTGYCCNACNGMNGMNCNATH 213
 QY 42 ProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValITrpglu 61
 DB 214 CCGGTATATGARGGNGAYGAYATHTGARGTNTTGNATTCARGCCNCGNTNGGAR 273
 QY 62 PheIysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArg 81
 DB 274 TTAAARAYAGNGAYTNTYNGNCAYTTAAARHTATGACAYGCGNATHTGNTTGMN 333
 QY 82 SerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsn 101
 DB 334 WSNACNTYTAACNGNAAARAYTAACNATGARGGARGAYGARYNTTTCARYYTNGNAAAY 393
 QY 102 CysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAla 121

DB 394 TGYACNTTCCNCAAYTTMNGCNCNGARATGAGYGCNCNTTGTGTGAAYCARGNGCN 453
 QY 122 AlaCysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuValGln 141
 DB 454 GCNTGYTTTYYTGARGNATGAAYGATNCAYTGGAAARGAAAYGNAACNTTNGTCNR 513
 QY 142 ValAlaThrIleSerGlyAspMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsn 161
 DB 514 GTTNGACNATHTMSNGNAAATATGTTTAAATCARATGCGNAARTGGGTAAARCARAGAYAY 573
 QY 162 GluThrGlyIleTyrTyrGluThrTrpAsnValLysAspSerProGluLysGlyAlaGlu 181
 DB 574 GARAACNGNATHTATATGACACNTGAAAYGTNARCGNWSGCCGAAARAGNGCGAR 633
 QY 182 ThrTrpPheAspSerTyrAspCysSerLysPheValIleuArgThrPheAsnLysLeuAla 201
 DB 634 ACNTGGTGYGAYMSNTAAGAYTGYMSNARTTYGNTTNGNACNTTAAATAYARVTNGCN 693
 QY 202 GIUPEGLYALAGLUPEHELYASNLILEGLUThrAsnTyrThrArgIlePheLeuTyrSer 221
 DB 694 GARTTGGNGCNGARTTAAARAYATHTGARACNAAATAYACMGNAATHTTYYTNTAYMSN 753
 QY 222 GlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThr 241
 DB 754 GGNGARCCNACNTAYTTNGNAAAYGARACMWSNGNTTGGCCNACNGNAAAYARACN 813
 QY 242 LeuGlyLeuAlaIleLysArgPheTyrTrpProPheLysProHisLeuProThrLysGlu 261
 DB 814 YTMGNNTYNGCATTAARAGNTTYTAYTAYCCNTTAAACNCAAYTTCNACNARARG 873
 QY 262 PheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeu 281
 DB 874 TTYTNTYTMWSNTYNTNCARATHTTGYAGCNGNATHTGNCAYARCAFTTAYTYN 933
 QY 282 PheTyrAsnPheGluTyrTrpPheLeuProMetLysPheProPheIleLysIleThrTyr 301
 DB 934 TTYTAAATTTGARAYAGTGYTTCNATGAGARTTYCCNTTATTAARATHTACNTAY 993
 QY 302 GIUGINTLEPROLEUPROILEARGASNLVSTHLEUSERGLYLEU 315
 DB 994 GARGARATHTCCNYTNCNATHTMGNAAAYARACNTYTMWSNGN 1035
 RESULT 6
 US-10-010-050A-13
 ; Sequence 13, Application US/10010050A
 ; Publication No. US20020173624A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; APPLICANT: Gilbertson, Debra G.
 ; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
 ; FILE REFERENCE: 97-38C1
 ; CURRENT APPLICATION NUMBER: US/10/010,050A
 ; PRIOR FILING DATE: 2002-03-26
 ; PRIOR APPLICATION NUMBER: US 09/122,383
 ; PRIOR FILING DATE: 1998-07-24
 ; PRIOR APPLICATION NUMBER: US 60/053,613
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 1038
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Degenerate nucleotide sequence encoding zs1946
 ; OTHER INFORMATION: polypeptide of SEQ ID NO:2
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1038)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-10-010-050A-13

Alignment Scores:

Pred. No.:	2,276-190	Length:	1038
Score:	1515.00	Matches:	267
Percent Similarity:	85.03%	Conservative:	0
Best Local Similarity:	85.03%	Mismatches:	47
Query Match:	86.18%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050a-2_COPY_31_346 (1-316) x US-10-010-050a-13 (1-1038)

```

QY      2 ValSerGlyLeuProSerArgGhiStrProValProTyrLysArgPheAspPheArg 21
DB      |||
QY      94 GTWMSGNGAATGCCNSMNGMNGCAYTGCCGNGCNGCTAYAAAGMNTTGYATTTMG 153
QY      22 ProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerProile 41
DB      |||
QY      154 CCAAAACCGACATCCTAATTCYCARCGMAATTAACCTTTTGCCACACGNGMNCNAH 213
QY      42 ProValMetGluLysAspAspAspIleGluValPheArgLeuGlnAlaProValTTPGlu 61
DB      |||
QY      214 CCGTAAATGARGGNGAYGAYGAYATHGARGTNTTVMGNTTNCARGCNGCTGTGGAR 273
QY      62 PheLysTyrGlyAspLeuLeuGlyLysLeuLysIleMetHisAspAlaIleGlyPheArg 61
DB      |||
QY      274 TTYAARATAYGNGAYTNTYNTGNCAYTTAAARATATATGCAAYGCMATHGENTTYMG 333
QY      82 SerThrLeuThrGlyLysAsnTyrThrMetGluTyrGlyLeuLeuPheGlnLeuGlyAsn 101
DB      |||
QY      334 WSAACNTYNAACGNAARAAATTAACATGAGATGTAATGATTTTTCARNTNGNAAY 193
QY      102 CysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAla 121
DB      |||
QY      394 TGYACNTTCCACAYTNTMNGCNGARATGAGCNGCCTTGYGTGYAAYCARGNGNG 453
QY      122 ALcAspPhePheGluGlyIleAspAspValHisTrrLysGluAsnGlyThrLeuValGln 141
DB      |||
QY      454 GCMTGTTTTTTCARCGNATHGAYGAYGNTCAATGGAARARAAAYGNAACNTYNTGNCAR 513
QY      142 ValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTrrValLysGlnAspAsn 161
DB      |||
QY      514 GTGCGACACATHTMSNGNAAYATGTTTAAACATGCGNAAATGGGTMAACRGAAAY 573
QY      162 GluThrGlyLleTyrTyrGluThrTTPAsnValLysAspProGluLysGlyAlaGlu 181
DB      |||
QY      574 GARAACNGMNTTAYTAYGARAACNTGGAAYGTNAAGCMNSMNCNGAARAAAGNGCNGAR 633
QY      182 ThrTrrPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeuAla 201
DB      |||
QY      634 ACNTGGTTTTCAYMSNTAYGAYTGYMSNAATTTGTYTNNGACNTTAAAYAAAYTNGCN 693
QY      202 GluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSer 221
DB      |||
QY      694 GARTTYGNGCNGARTTAAARAAVAITHGARAACNAAYAAVACMNGMNTHTTNTATWSN 753
QY      222 GlyLysProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThr 241
DB      |||
QY      754 GAGNARCCNACNTAYTNTGNAAYGARAACWSMNTTNTTGGCNCNACNGNAAYAAARACN 813
QY      242 LeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLysGlu 261
DB      |||
QY      814 YTTNGNTTNGCNAATHAARMGNTTAYATAYACCTTAAACCCNAYTNTCCNACNAAGAR 873
QY      262 PheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeu 281
DB      |||
QY      874 TTYTNTTNNSTNTYNTNCAATHTTTCAGGNGTNTATGNCVAVARCARCTTAYTNT 933
QY      282 PheTyrAsnPheGlyLysTrrPheLeuProMetLysPheProPheIleLysIleThrTyr 301
DB      |||
QY      934 TTYTAAVAYTTCGARTYATGTTTNTCCNATGARTTTCNTTAAATHAARATACNTAY 993
QY      302 GluGluLeuProLeuProLeuArgAsnLysThrLeuSerGly 315
DB      |||

```

DB 994 GARGARATHCCNTYNTCCNATHTMGNAAYAAARACNTTMSNGN 1035

RESULT 7

```

US-09-864-761-22592
/ Sequence 22592, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Mensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aecmlca-X-1
/ CURRENT FILING DATE: US/09/864,761
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR FILING DATE: US 60/207,456
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annonak Sequence Listing Engine vers. 1.1
/ SEQ ID NO 22592
/ LENGTH: 494
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC001226.1
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
/ OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
/ OTHER INFORMATION: NT HIT: g45729771, EVALU0.00e+00
/ OTHER INFORMATION: SWISSPROT HIT: 075503, EVALU0.00e+00
/ OTHER INFORMATION: EST_HUMAN HIT: H01255.1, EVALU0.00e+00
US-09-864-761-22592

```

Alignment Scores:

Pred. No.: 2,07e-107 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 50.40% Indels: 0
 DB: 9 Gaps: 0

US-10-010-050a-2_COPY_31_346 (1-316) x US-09-864-761-22592 (1-494)

QY 154 AAlaYsTPVallyVgInAspAsnGluThrsGlylleYrYrGluThrTPAsnVallye 173
 Db 3 GCAAGTGGTGAACAGGACATGAAACAGAAATTATATGACATGAAAGTAAAA 62
 QY 174 AAlaserProGluYrGlyAlaGluThrTPheAspSerTyraSPCySerTyraPheVal 193
 Db 63 GCCAGCCCAAGAAAGGGGGCAGAGACATGCTTATCTTACGACTGTTCCAAATTTGTG 122
 QY 194 LeuArGrThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnLleGluThrAsn 213
 Db 123 TTAAAGACCTTTAAACAGTGGCTGATTTGGAGCAGAGTTCAAGAACATGAAACCAAC 182
 QY 214 TyrThrArgIlePheLeuTyraSPGlyGluProThrTyraSPGlyAsnGluThrSerVal 233
 Db 183 TATACAGAAATATTTCTTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTGTT 242
 QY 234 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyraSPhe 253
 Db 243 TTTGGGCCCAAGAGAAACAGACCTTGCTTACCAATTTATTTATCCCTTC 302
 QY 254 LysProHisLeuProThrLysGluPheLeuSerLeuGluIlePheAspAlaVal 273
 Db 303 AAACCACTTTGCCACTAAAGAAATTTCTGTGAGTCTCTTGCAGAAATTTGATGACGTG 362
 QY 274 IleValHisLysGlnPheTyraSPheTyraSPheGlyTyraSPheLeuProMetLys 293
 Db 363 ATGTGCACAAACAGTCTTATTTGTTATATTTGAAATATGCTTTTACCATGAAA 422
 QY 294 PheProHelleLysIleThrTyraSPGluGluIleProLeuProIleAsnLysThrLeu 313
 Db 423 TTCCTTTTATTAATAATACATATGAAAGAAATCCCTTACCTATCAGAAACAAACACTC 482
 QY 314 SerGlyLeu 316
 Db 483 TCTGTTTA 491

RESULT 8

US-10-040-739-825
 Sequence 825, Application US/10040739
 Publication No. US20020173635A1

GENERAL INFORMATION:

APPLICANT: McCoy, Kenneth
 McCoy, John
 Lavallee, Edward
 Racine, Lisa
 Merberg, David
 Treacy, Maurice
 Spaulding, Vikki
 TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
 NUMBER OF SEQUENCES: 1519
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy Disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/040,739
 FILING DATE: 07-Jan-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/036,520
 FILING DATE: 03-JUN-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Brown, Scott A.
 REGISTRATION NUMBER: 32,724
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 498-8224
 TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 825:

SEQUENCE CHARACTERISTICS:
 LENGTH: 506 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 825:

US-10-040-739-825

Alignment Scores:

Pred. No.: 2,73e-86 Length: 506
 Score: 727.00 Matches: 133
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 41.35% Indels: 0
 DB: 14 Gaps: 0

US-10-010-050a-2_COPY_31_346 (1-316) x US-10-040-739-825 (1-506)

QY 121 AAlaCyPhePheGluGlylleAspAsnValHisTPlyGluAsnGlyThrLeuVal 140
 Db 99 GCTGCCCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 158
 QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGluMetAlaIleSTPVallyGluAsp 160
 Db 159 CAAGTGCACATATATGAGAAACATGTTCAACCAATGCGAAAGTGGTGAACAGAGAC 218
 QY 161 AsnGluThrsGlylleYrYrGluThrTPAsnVallyAsnLaserProGluYrGlyAla 180
 Db 219 AATGAAACAGGAATTTATATGACATGGAATGTAAACCAACCCAGAAAGGGGCA 278
 QY 181 GluThrTPheAspSerTyraSPCySerTyraPheValleuArGrThrPheAsnLysLeu 200
 Db 279 GAGACATGTTGATTCCTACGACGCTGTTCCAAATTTGTAAAGACCTTTAAACAGTTG 338
 QY 201 AAlaGluPheGlyAlaGluPheLysAsnLleGluThrAsnTyraSPheLeuTyra 220
 Db 339 GCTGAATTTGAGCAGAGTTCAAGAACATGAAACCAACTATACAGAAATATTTCTTAC 398
 QY 221 SerGlyLysProThrTyraSPGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
 Db 399 AGTGGAGAACCTTATCTGCGGAATGAAACATCTGTTTGTGGCCACAGGAAACAG 458
 QY 241 ThrLeuGlyLeuAlaIleLysArgPheTyraSPhe 253
 Db 459 ACTCTTGTTTACCATTAATAAGATTTATTAACCCCTTC 497

RESULT 9

US-09-864-761-5832
 Sequence 5832, Application US/09864761
 Patent No. US2002048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wenheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aeomica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761

```

CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 5832
LENGTH: 474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC001226.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
US-09-864-761-5832

Alignment Scores:
Pred. No.: 1,07e-80      Length: 474
Score: 684.50           Matches: 129
Percent Similarity: 86.67%      Conservative: 1
Best Local Similarity: 86.00%    Mismatches: 1
Query Match: 38.94%           Indels: 19
DB: 9                      Gaps: 1

US-10-010-050a-2_copy_31_346 (1-316) X US-09-864-761-5832 (1-474)

QY 123 CysPhePheGluGluY1LeaspAspValHsTrpLysGluAsnGlyThrLeuValGlnVal 142
DB 82 TCTTTTAAAA----- 93
QY 143 AlaThrIleSerGlyAsnMetPheAsnGlnMetAlaValTrpValLysGlnAspAsnGlu 162
DB 94 -----CTAGGAACATGTTTCAACCAATGCGAAAGTGGGTGTAACGACACATGAA 144

```

```

QY 163 ThrGlyIleTyTrpGluThrTrpAsnValLysAlaSerProGluLysGlyValGluThr 182
DB 145 ACAGGAATTTATTTATGACACATGCAATCTAAAGCCAGCCAGAAAAGGGCCAGACACA 204
QY 183 TrpPheAspSerTyTrpAspCysSerLysPheValLeuArgThrPheAsnLysLeuAlaGlu 202
DB 205 TGGTTTGATCTCTCAAGCATGCTTCCAAATTTGTGTTAAGACCTTTAAACAAGTGGCTGA 264
QY 203 PheGlyAlaGluPheLysAsnLleGluThrArgTrpThrArgLlePheLeuTySerGly 222
DB 265 TTGGAGCAGAGGTCACAGAAACATAGAAACCAACTATACAAATAATTTCTTTAAGATGA 324
QY 223 GluProThrTyTrpLysGluAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeu 242
DB 325 GAACCTACTTATCTGGGAAATGAAACATCTGTTTGGGCCAACAGAAACAAAGCTCTT 384
QY 243 GlyLeuAlaIleLysArgPheTyTrpProPheLysProHsLeuProThrLysGluPhe 262
DB 385 GGTWTAGCCATTAATAAGATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 444
QY 263 LeuLeuSerLeuLeuGlnIlePheAspAla 272
DB 445 CTGTGAGTCTCTGCAAAATTTTGATGCA 474

RESULT 10
US-10-264-049-1859
Sequence 1859, Application US/10264049
Publication No. US20040005579A1
GENERAL INFORMATION:
APPLICANT: Birex et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA133P1
CURRENT APPLICATION NUMBER: US/10/264,049
CURRENT FILING DATE: 2002-10-04
PRIOR APPLICATION NUMBER: PCT/US01/18569
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: US 60/209,467
PRIOR FILING DATE: 2000-06-07
NUMBER OF SEQ ID NOS: 4360
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 1859
LENGTH: 497
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (2)..(2)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (175)..(175)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (285)..(285)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (295)..(295)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (313)..(313)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (323)..(324)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (331)..(331)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:

```

```
/ NAME/KEY: misc_feature
/ LOCATION: (395)..(395)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (440)..(440)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (458)..(458)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (464)..(464)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (478)..(478)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (483)..(483)
/ OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-1859
```

```
Alignment Scores:
Pred. No.: 1.85e-52 Length: 497
Score: 472.00 Matches: 88
Percent Similarity: 98.88% Conservative: 0
Best Local Similarity: 98.88% Mismatches: 1
Query Match: 26.85% Indels: 0
DB: 16 Gaps: 0
```

US-10-010-050a-2_COPY_31_346 (1-316) x US-10-264-049-1859 (1-497)

```
QY 228 GYANGIURHRSERVALPHEGIYPROTHGIYASNLVSTHLEUGLYLEUALALELYS 247
DB 3 GGAATGAAACATCTGTTTGGGCCAAGGAAACAGCTCTGTTTGAAGCTCTCTG 52
QY 248 ARGPHETRYRPRORPELYSPROHISLEUPROTHILYSGIUPHELEUSERLEU 267
DB 63 AGATTTATTAACCCCTTCAACACATTTGCCAATAAGATTTCTGTGAGCTCTCTG 122
QY 268 GNIIEPHEASPAIALVALIIEVALIISLYSGINPHEIYRLEUPHETRYRANPHEIUTYR 287
DB 123 CAATTTTGTATGAGCATTTGTGCACAAACAGTTCTATTGTTTAAATTTGAATAT 182
QY 288 TRPHELEUPROECLYSPPHEIOPHEIILEYSLIETHTYRGIUGIUILEPROLEUPRO 307
DB 183 TGGTTTTCCTATGAAATTCCTTTTATTAATAATACATATGAGAAATCCCTTACT 242
QY 308 ILEARGASNLVSTHLEUSERGIYLEU 316
DB 243 ATCAGAAACAAACACTCTGTGTTTA 269
```

RESULT 11

```
US-09-867-701-2579
/ Sequence 2579, Application US/09867701
/ Patent No. US20020132237A1
/ GENERAL INFORMATION:
/ APPLICANT: Agiate, Paul A.
/ APPLICANT: Jones, Robert
/ APPLICANT: Harlocker, Susan L.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
/ FILE REFERENCE: 210121.497
/ CURRENT APPLICATION NUMBER: US/09/867, 701
/ NUMBER OF SEQ ID NOS: 10912
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2579
/ LENGTH: 351
/ TYPE: DNA
```

ORGANISM: Homo sapien
US-09-867-701-2579

```
Alignment Scores:
Pred. No.: 9e-38 Length: 351
Score: 359.50 Matches: 78
Percent Similarity: 68.25% Conservative: 8
Best Local Similarity: 61.90% Mismatches: 25
Query Match: 20.45% Indels: 16
DB: 9 Gaps: 3
```

US-10-010-050a-2_COPY_31_346 (1-316) x US-09-867-701-2579 (1-351)

```
QY 82 SerThrlEuthrGIYLYSASNLVSTHMEGLUTRPIYRGIULEPHEGLNLEUGLYASN 101
DB 3 AGTACATTAAGTGCAGAACTACACAAATGAAATGATATGACTTTCCACTTGGCAAC 62
QY 102 CysThrPheProHISLEUAHGTGPROGLUMETASPAALAPROPHETPCYSASNLINLYA 121
DB 63 TGTACATTTCCCGCATCTCCGACCTGAAATGATGATCCCTTCTGTGTATCAAGCGCT 122
QY 122 AlaCysPhePheGluGIYIIEASPAAPVALHISIRPLYSGLUASNLVSTHLEUVALGLN 141
DB 123 GC-TGCTTTTGTAGGGAATTTGATGATGTCACCTGAAAGAAATGGACATTTAGTCAA 181
QY 142 ValAlaThrIleSerGIYASNMet-----PheASNLINMETALALYSTRPVALYSGLN 159
DB 182 GTAGCAACTATATACAGTAAAGTTGTGTAATAATATGCAATTTGATGATGCACCAAAA 241
QY 160 AspASNLINHTRGIYIETRYTYRGIUTHTRPASNVALYBALASERPROGLULYSGLY 179
DB 242 CCAATGAAAGAAATTTGTAT-----ACTTCC 268
QY 180 AlaGIUTHTRPHEAPSERIYRASPYSERLYPHEVALLEUARGTHRPASN--- 198
DB 269 ATGGAACATTTACGATATGTTTACTTACAGGCTCATTTGTAATGATCTTTGGAAC 328
QY 199 ---LysLEUALAGLUPE 203
DB 329 ATTAACCTTTGGCAATTT 346
```

RESULT 12

```
US-09-954-456-1363/C
/ Sequence 1363, Application US/09954456
/ Patent No. US20020115057A1
/ GENERAL INFORMATION:
/ APPLICANT: Young, Paul
/ TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cance
/ TITLE OF INVENTION: Sers
/ FILE REFERENCE: 689290-76
/ CURRENT APPLICATION NUMBER: US/09/954,456
/ PRIOR FILING DATE: 2001-09-18
/ PRIOR APPLICATION NUMBER: US/60/233,617
/ PRIOR FILING DATE: 2000-09-18
/ PRIOR APPLICATION NUMBER: US/60/234,052
/ PRIOR FILING DATE: 2000-09-20
/ PRIOR APPLICATION NUMBER: US/60/234,923
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US/60/235,134
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US/60/235,637
/ PRIOR FILING DATE: 2000-09-26
/ PRIOR APPLICATION NUMBER: US/60/235,638
/ PRIOR FILING DATE: 2000-09-26
/ PRIOR APPLICATION NUMBER: US/60/235,711
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: US/60/235,720
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: US/60/235,840
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: US/60/235,863
/ PRIOR FILING DATE: 2000-09-27
/ NUMBER OF SEQ ID NOS: 2276
```

```
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 1363
/ LENGTH: 473
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-954-456-1363

Alignment Scores:
Pred. No.: 7,81e-22 Length: 473
Score: 241.00 Matches: 44
Percent Similarity: 100.00% Conservative: 44
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13,71% Indels: 0
DB: 9 Gaps: 0

US-10-010-050A-2_COPY_31_346 (1-316) x US-09-954-456-1363 (1-473)

QY 273 ValIleValHisGlnGlnPheTyrrLeuPheTyrrAsnPhgIuTyrrTrpPheLeuPromet 292
DB 472 GTGATGTGCACAAAGCTTCATTTGTTTATTAATTTGAAATATGTTTACCTATG 413
QY 293 LysPheProPheIleuLysIleThyTyrGluGluIleProLeuProIleArgAsnLysThr 312
DB 412 AATTCCTCTTTATTAATAATGAAATGAAATCCCTTACTTACGAAACAAACA 353
QY 313 LeuSerGlyLeu 316
DB 352 CTCTCTGGTTTA 341

RESULT 13
US-09-908-975-10339
/ Sequence 10339, Application US/09908975
/ Publication No. US20030165843A1
/ GENERAL INFORMATION:
/ APPLICANT: SHOSHAN, Avi
/ APPLICANT: WASSERMAN, Alon
/ APPLICANT: MINTZ, Eli
/ APPLICANT: MINTZ, Liat
/ APPLICANT: FAIGER, Simchon
/ TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC
/ FILE REFERENCE: 36688-0005
/ CURRENT APPLICATION NUMBER: US/09/908,975
/ PRIOR FILING DATE: 2001-07-20
/ PRIOR APPLICATION NUMBER: US 60/287,724
/ PRIOR FILING DATE: 2001-05-02
/ PRIOR APPLICATION NUMBER: US 60/221,607
/ NUMBER OF SEQ ID NOS: 32337
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 10339
/ LENGTH: 60
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-908-975-10339

Alignment Scores:
Pred. No.: 2,3e-07 Length: 60
Score: 121.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6,88% Indels: 0
DB: 10 Gaps: 0

US-10-010-050A-2_COPY_31_346 (1-316) x US-09-908-975-10339 (1-60)

QY 102 CysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
DB 3 TGTACATTTCCCATCTCCGACCTGAATAATGATGCCCTTCTGTGTATCAAGCC 59

RESULT 14
US-10-198-846-9753/C
/ Sequence 9753, Application US/10198846
```

```
/ Publication No. US20030099974A1
/ GENERAL INFORMATION:
/ APPLICANT: Lillie, James
/ APPLICANT: Xu, Yongyao
/ APPLICANT: Wang, Youzhen
/ APPLICANT: Steinmann, Kathleen
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
/ TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
/ FILE REFERENCE: MRI-049
/ CURRENT APPLICATION NUMBER: US/10/198,846
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/306,220
/ NUMBER OF SEQ ID NOS: 14084
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 9753
/ LENGTH: 2667
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 2667
/ OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9753

Alignment Scores:
Pred. No.: 0.017 Length: 2667
Score: 104.50 Matches: 70
Percent Similarity: 30.72% Conservative: 32
Best Local Similarity: 21.08% Mismatches: 109
Query Match: 5,94% Indels: 121
DB: 15 Gaps: 15

US-10-010-050A-2_COPY_31_346 (1-316) x US-10-198-846-9753 (1-2667)

QY 22 ProLysProAspProTyrCysGlnAlaLysTyrThrPhe 34
DB 1674 CCTGCGCCCGGGGCTCTTGTGCACACACCCAGGCTTCCAGAAACAGAGACCGGG 1615
QY 35 38
DB 1614 AGCTGACAGTGAGAGACACCTCTATGCGTGTGCCACCTTCCACATGAGTGAGG 1555
QY 39 51
DB 1554 CAACATGACAGAGCCCAACAGCTGAGCCATCTCTCACTGAGCCTGAGAGGATCCAC 1495
QY 52 ValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAspLeuLeuGlyHisLeu 71
DB 1494 GTAATGTGTCTTCAAGTCCCG---TGGGAATG---TGGCATCTC 1456
QY 72 LysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThrMet 91
DB 1455 AGCCTTCCAGAGGTGTGGCACTCACGCTCCACATCTCCCAACAA---ACGCNA 1402
QY 92 GluTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGlu 110
DB 1401 GGGTCTCTGCTCTTACTAATA---AACCTTCCCTTCCAGAAACAGCCCAACTC 1345
QY 111 115
DB 1344 ATCACTGTTCAATATGTCGCAAGTGAGGCTCTCTGCTCCCTTTACCAAGAGACGGA 1285
QY 116 TrpCysAsnGlnGlyAlaIleCysPhePheGlnGly 127
DB 1284 TGGTGTCTGCAAGGCAAGTCTCTGAGTGTGAAGATGAGCCCTCAGGCTCCCAAC 1225
QY 128 145
DB 1224 TGGCAATACAGAGCTGTGAGCGGCTGGAGAGGAGTCTCTTCTCTCCATCCAGCGG 1165
QY 145 LeuSerGlyAsnMetPheAsnGlnMetAlaLys---TrpValLysG 159
```

```

Db      1164 TCACCGGATCTCTCCGGGAGAAAGCCAAACAGCCGCTTCCCTTCGG-----T 1111
QY      159  InaPaangluThnGlyIleTyrtYrGluthrTrpAenValLysAlaSerProGluLysG 179
Db      1110 CACGACGAGCCAGGCGGTTCCAGCAGAGGCTGTGGCAACATGTCATCCGCCCAT--G 1054
QY      179  lYalaGluthrTrpPheAspSerTyraSPCySerLysPheValLeuArgThrPheAsnL 99
Db      1053 GTGMAATGGGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1001
QY      199  yLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrtYrThrArgIlePheL 219
Db      1001 ----- 1001
QY      219  eutyrserygluProthrTyrlenglyasnGluThrSerValPheGlyProthGlyA 239
Db      1000 -----GCTGGAC 994
QY      239  snLysThrleuGlyLeuAlaIleLysArgPheTyrtYrProPheLysProHISLeuPro 259
Db      993  CCCTGCTCTGGGCGGACCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 934
QY      259  hrlYsgLuphe-----LeuIeuSerLeuLeuGlnIlePheAspAlaValIleV 275
Db      933  GCCAGTAATTTCTCTCTTTACTTCTGCGGCTTCTCGGCGCTTGGCGTCTGCTCTCC 674
QY      275  aHISLysGlnPheTyrlEuPheTyraSnPhe 285
Db      873  TCTGCAAGCATTTATTCTTTCTTCAGTCTTTT 842

```

RESULT 15

```

US-10-369-493-36401
; Sequence 36401, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 36401
; LENGTH: 986
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-36401

```

Alignment Scores:

```

Pred. No.: 0.0287 Length: 986
Score: 97.50 Matches: 41
Percent Similarity: 38.22% Conservative: 19
Best Local Similarity: 26.11% Mismatches: 50
Query Match: 5.55% Indels: 47
DB: 16 Gaps: 7

```

JS-10-010-050a-2_COPY_31_346 (1-316) x US-10-369-493-36401 (1-986)

```

JY      138  ThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAla----- 154
Jb      168  ACGTTGGTGTGATCGCGGCTTCCAAAGCAACACACGCGGCAAGTGGCGGCGTGGCG 227
JY      155  -----LysTrpValLysGlnAspAsnGlu 162
Jb      228  CGAAGCTTGGCCTGGAAGTTGCGCTTGATACAGAAATAAGGTGATGGAGCAGCAAA 287

```

```

QY      163  ThrGlyIleTyrtYrGluThrTrpAenValLysAlaSerProGluLysGlyAlaGluThr 182
Db      288  AATGGCGTTTACGACAGAGCGGGAATATCCAGCTTTCGGCGCTGATGGCGCAGACAGC 347
QY      183  TrpPheAspSerTyraSPCySerLysPheValLeuArgThrPheAsnLysLeuAlaGlu 202
Db      348  AGACTCGAT-----ATGTCGCGC 365
QY      203  PheGlyAlaGlu-----PheLysAsnIleGluThrAsnTyrtYrThrArgIlePhe 218
Db      366  TTCGGATTGAACATTAATTGACCCCTCAGAACTTGGAG-----GAGAAATCAGA 416
QY      219  LeuTyrserygluProthrTyrlenglyasnGluThrSerValPheGlyProthGly 238
Db      417  GCCAGGCGGCAAGCCATATTACATC-----CCTGACGAGGATCATGAT 461
QY      239  AsnLysThrleuGlyLeuAlaIleLysArgPheTyrtYrProPheLysProHISLeuPro 258
Db      462  CACCCATTGGCGGCTTAGGCTTCCGACG-----TGGCGCTTGGAGTTGAGATGACAG 515
QY      259  ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleVal 275
Db      516  GAGAAAGAG-----CTCGGTGATTTCTTGAACACTGTGATTTGTG 554

```

Search completed: April 25, 2004, 06:14:09
Job time : 636.806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 23:02:52 ; Search time 104.894 Seconds
(without alignments)
1671.819 Million cell updates/sec

Title: US-10-010-050a-2_COPY_31_346
Sequence: 1 RVSGIPSRHMPVYKRFDF.....IKTYEELPIPNKTIISGL 316

Scoring table: BIOSUM62
Xgapop 10.0, Ygapext 0.5
Xgapop 10.0, Ygapext 0.5
Xgapop 6.0, Ygapext 7.0
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DRV=xjh
-O=/cgm2_1/USFTO.spool/US10010050/runat.22042004_113205_27569/app_query.fast1_1.1372
-DB=Issued_Patents_NA -OPMT=fastap -SUFTX=rni -MINMATCH=0.1 -IOOPCL=0
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10010050 @CGN 1.1 140 @runat.22042004_113205_27569 -NCPU=6 -ICPU=3
-NO MMAP -LARGEJOB -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents NA:*
1: /cgm2_6/pcodata/2/ina/5A.COMB.seq:*
2: /cgm2_6/pcodata/2/ina/5B.COMB.seq:*
3: /cgm2_6/pcodata/2/ina/5A.COMB.seq:*
4: /cgm2_6/pcodata/2/ina/5B.COMB.seq:*
5: /cgm2_6/pcodata/2/ina/PCUTS.COMB.seq:*
6: /cgm2_6/pcodata/2/ina/backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	5.4	14070	4	US-09-108-006C-2
2	91.5	5.2	7718	4	US-09-976-594-244
3	90.5	5.1	1349	1	US-08-290-448A-73
4	90.5	5.1	1349	1	US-08-290-448A-73
5	90.5	5.1	1349	1	US-08-290-448A-73
6	90.5	5.1	1349	1	US-08-175-069A-73
7	90.5	5.1	1349	4	US-08-461-939B-73
8	89.5	5.1	1328	1	US-08-464-000-73
9	89.5	5.1	1328	1	US-08-290-448A-58
10	89.5	5.1	1328	1	US-08-290-448A-58
11	89.5	5.1	1328	4	US-08-175-069A-58
12	89.5	5.1	1328	4	US-08-461-939B-58
					Sequence 58, Appl

13	88.5	5.0	4440	3	US-07-792-600-1	Sequence 1, Appl
14	88.5	5.0	4440	3	US-09-157-021-1	Sequence 1, Appl
15	88.5	5.0	4440	3	US-09-156-842-1	Sequence 1, Appl
16	88.5	5.0	4440	4	US-09-591-514-1	Sequence 1, Appl
17	88.5	5.0	5433	3	US-09-157-021-35	Sequence 35, Appl
18	88.5	5.0	5433	3	US-09-156-842-35	Sequence 35, Appl
19	88.5	5.0	5433	4	US-08-023-655-1363	Sequence 35, Appl
20	88.5	5.0	5433	4	US-09-591-514-35	Sequence 35, Appl
21	88.5	5.0	5433	4	US-09-620-312D-253	Sequence 35, Appl
22	88.5	5.0	1827	4	US-09-489-039A-6332	Sequence 22, Appl
23	87.5	4.9	13930	4	US-09-134-000C-219	Sequence 21, Appl
24	86.5	4.9	13930	4	US-09-976-594-1011	Sequence 21, Appl
25	86.5	4.9	13930	4	US-09-220-132-20	Sequence 21, Appl
26	85.5	4.9	3155	2	US-08-591-629-7	Sequence 20, Appl
27	85	4.8	1830121	4	US-09-557-884-1	Sequence 1, Appl
28	85	4.8	1830121	4	US-09-643-990A-1	Sequence 1, Appl
29	84	4.8	3629	1	US-08-354-618-1	Sequence 1, Appl
30	83	4.7	1116	4	US-09-252-991A-3059	Sequence 1, Appl
31	83	4.7	1668	4	US-09-134-078-10	Sequence 3059, Ap
32	83	4.7	9139	4	US-09-332-478-22	Sequence 22, Appl
33	83	4.7	1664976	4	US-08-916-421B-1	Sequence 1, Appl
34	81.5	4.6	1332	4	US-09-540-236-1387	Sequence 1387, Ap
35	81	4.6	2226	4	US-09-489-039A-683	Sequence 683, Ap
36	80.5	4.6	2157	4	US-09-328-352-514	Sequence 514, Ap
37	80.5	4.6	4643	4	US-09-453-702B-22	Sequence 22, Appl
38	80	4.6	1356	4	US-09-489-039A-4678	Sequence 4678, Ap
39	80	4.6	1821	4	US-09-537-682-2	Sequence 2, Appl
40	80	4.6	1959	4	US-09-489-039A-5290	Sequence 5290, Ap
41	79.5	4.5	507	3	US-08-384-106A-3	Sequence 3, Appl
42	79.5	4.5	507	5	PCT-US96-01643-3	Sequence 3, Appl
43	79.5	4.5	594	4	US-09-489-039A-1146	Sequence 1146, Ap
44	79.5	4.5	694	2	US-08-627-610-9	Sequence 9, Appl
45	79.5	4.5	694	3	US-08-384-106A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-108-006C-2
; Sequence 2, Application US/09108006C
; Patent No. 6524613
; GENERAL INFORMATION:
; APPLICANT: Steer, Clifford J.
; Kren, Betsy T.
; Bandyopadhyay, Paramita
; Roy-Chowdhury, Jayanta
; TITLE OF INVENTION: Hepatocellular Chimeraplasty
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kimeragen, Inc.
; STREET: 300 Pheasant Run
; CITY: Newtown
; STATE: PA
; COUNTRY: USA
; ZIP: 18940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/108, 006C
; FILING DATE: 30-Jun-1992
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/054,288
; FILING DATE: 30-APR-1997
; APPLICATION NUMBER: 60/054,837
; FILING DATE: 05-AUG-1997
; APPLICATION NUMBER: 60/064,996
; FILING DATE: 10-NOV-1997
; APPLICATION NUMBER: 60/074,497
; FILING DATE: 12-FEB-1998

```

APPLICATION NUMBER: PCT US 98/08834
FILING DATE: 30-APR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Filebel, Thomas
REGISTRATION NUMBER: 29258
REFERENCE/DOCKET NUMBER: 7991-015-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-504-4444
TELEFAX: 215-504-4545
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 14070 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-108-006C-2

Alignment Scores:
Pred. No.: 1.44 Length: 14070
Percent Similarity: 94.50 Matches: 65
Percent Similarity: 34.57% Conservative: 47
Best Local Similarity: 20.06% Mismatches: 122
Query Match: 5.38% Indels: 90
DB: 4 Gaps: 13

US-10-010-050A-2_COPY_31_346 (1-316) x US-09-108-006C-2 (1-14070)
QY 2 ValSerGlyIleProSerArgArgHisIleThrProValProTyrLysArgPheAsp-PheAr 21
DB 12075 CTCACGACCTCCCGACCGGAGGAC--CGGGCGATGATAT 12116
QY 21 gProLysProAspProTyrCySGlnAlaLysTyrThrPheCysProThrGlySerPro1 41
DB 12117 GATGATAGATGACGACCTTTCTTAATGAGACTTCTACTACAGCCCTCTCTCCA-- 12174
QY 41 eProValMetGlnGlyAspAspAspIleGluValPheArgLeuGlnAlaProVal1TrpG1 61
DB 12175 -----GATTAATAAACTCCACCATATTCATAAAGTGAAGGTCGCGGA 12218
QY 61 uPheLysTyrGlyAspLeuLeuGlnHisLeuLysIleMetHisAspAlaIleGlyPheAr 81
DB 12219 A-----TCTGATGAGAACTCAGATCAAGTT----- 12246
QY 81 gSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAs 101
DB 12247 -----AATTGGGAAGAAGAGAGCTTCTGCTT 12275
QY 101 nCyThrPheProHisLeuArgProGluMetAspAlaProPheTrpCyAsnGlnGlyAl 121
DB 12276 GCTAAC-----TCTCTGAAGAACAACGTGCC-----AAGGCCAC 12311
QY 121 aAlaCysPhePheGlnGlyIleAspAspValHisTrpLysGlnAsnGly---ThrLeuVa 140
DB 12312 AGGGGCTCTTATATATATATATGCAACAAGTACCTGGGAACAACAAGGCTCAGCTGAG 12371
QY 140 lGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrValLysGlnAs 160
DB 12372 AGAAGTGTCTCAAGCTGAGAAGAAATCTGCAGACCAAGCTGAGTTCATCA-- 12429
QY 160 pAsnGlnThrGlyLeuTyrTyrGluThrTrpAsnValLysLeuAspProGluLysGlyAl 180
DB 12430 -----GGGGCCATTAGGGAATTTGATGATGACAGAGGTTCCAGAAAGAGGC 12479
QY 180 aGluThrTrpPheAspSerTyrAspCySerLysPheValLeuArgThrPheAsnLysLe 200
DB 12480 CAGTGGGACCACTGGGACCTTACCA----- 12504
QY 200 uAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn---TyrThrArgIlePheLe 219
DB 12505 -----GAGTGAAGAGCAAGAGCCCAAGATCTGTACACAGAACTGTGAC 12548

```

```

QY 219 uTyrSerGlyGluProThrTyrLeuGlyAsnGlnThrSerValPheGlyProThrGlyAs 239
DB 12549 TCAGGAAGCCCAAGCACTTCTCCAGGACTCAAGATTAACGTTTGATGCTGGTACG 12608
QY 239 nLysThrLeuGlyLeuAlaIleLys----- 247
DB 12609 AGTTACTCAGAAATTCATATGAAAGTCAGACATCTATGACTCATCATTTTCT 12668
QY 248 -----ArgPheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuLe 264
DB 12669 GAACCTCCCGAGATTCCAGTTCCGCGGAAACCTGGATATACACTAGGAGGAACCTTG 12728
QY 264 uSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyr----- 280
DB 12729 CACTATGTC---ATAGGAGGATGAGGACGATCTGCTCCAGGTATTCGAAAGTCCA 12785
QY 281 -----LeuPheTyrAsnPheGluTyrTrpPheLeuProMetLysPhePr 295
DB 12786 TAATGTTTCAAGAAATAGCTTTTCTTATTTCCAAAGACTAGTATTAACCTTCTTGA 12845
QY 295 oPheIleLys 298
DB 12846 GTTAAGGAA 12855

RESULT 2
US-09-976-594-244
Sequence 244, Application US/0976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
APPLICANT: Buchbinder, Jenny
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976,594
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 244
LENGTH: 7718
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Inocyte ID No. 6673549 322303.15
NAME/KEY: unsure
LOCATION: 6359-6362, 6967-7026
OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-244

Alignment Scores:
Pred. No.: 1.35 Length: 7718
Percent Similarity: 91.50 Matches: 71
Percent Similarity: 31.40% Conservative: 37
Best Local Similarity: 20.64% Mismatches: 133
Query Match: 5.20% Indels: 103
DB: 4 Gaps: 15

US-10-010-050A-2_COPY_31_346 (1-316) x US-09-976-594-244 (1-7718)
QY 11 TrpProValProTyrLysArgPheAsp-----PheArgProLysProAspProTyrCyS 28
DB 3443 TGGATGGCTCCGGAATCTATCTTTGACAAATCTACAGCAAGAGGACGATGCTCT 3502
QY 29 GlnAlaLysTyr-----ThrPheCyProThrGlySerProIleProValMetGlu 45
DB 3503 TACGAGTATTCGCTGGGAAATCTTCTCTTAAAGTGGGTCTCCATACCAAGAGTAA 3562
QY 46 GlyAspAspAspIle----- 50
DB 3563 ATGATGAGAACTTTTGACAGTCCCTGAGGAAAGGCAATGAGATGAGAGCTCTGAGTAC 3622

```

```

QY 51 -----GluValPheArgLeuGlnAlaProValTyrGlu-----61
Db 3623 TCTACTCTGAAATCATCATGACTGAGCTGCTGCGACAGACACCAAGAAAGAGG 682
QY 62 -----PheLysTyrGlyAspLeuLeuGlnHisLeuLysIleMet 74
Db 3683 CCAAGATTGACAGACTGTTGAGAAAACTAGTGATTGCTT---CAAGCAATGTACAA 739
QY 75 HisAspAlaIleLeuLysPheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyr 54
Db 3740 CAGAGT-----GGTAAAGACTACATCCCA-----ATC 3766
QY 95 GluLeuPheGlnLeuGlyAsnGlyThrPheProHisLeuArgProGluMetAspAlaPro 114
Db 3767 AATGCATATGACAGCAAGAAATAGTGGTTTACTACTCAACTCTCTCTCTCTGAGAC 3826
QY 115 PheTyrPheAsnGlnGlyAlaAlaCysPhePheGlu-----GlyIleAspAspValHis 132
Db 3827 TTCTTCAGAGAAAGTATTTCAAGCTCCAGATTAAATTCAGAAAGCTGTATGATGCAGA 3886
QY 133 TrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln 152
Db 3887 TATGTAATGCTTTCAAGTTCAATGAGCTCGAGAAAGATC-----AAAACTTTGAGAA 3940
QY 153 MetAlaLysTyrValLysGlnAspAsnGluTyrGlyIleTyr-----TyrGlu 168
Db 3941 CTTTAA-----CCGAATGCCACTCTCATCTTTGAGACACTACAGAGCGAC 3985
QY 169 ThrTyrPheAsnValLysLeuAspProGluLysGlyAlaGluThrTyrPheAspTyrAsp 188
Db 3986 AGCAGCACTCTGTGGCTCTCCCATGCGAGAGCGCTTCACTGAGACTGACAGCAAAACC 4045
QY 189 CysSer-----LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGly----- 204
Db 4046 AAGGCGCTCGCTCAAGATTGACTTTGAGAGTAAACAGTAAAGTAAAGTGGGCGCTGCT 4105
QY 205 -----AlaGluPheLysAsn 209
Db 4106 GATGTCAGCAGGCCCAAGTTTCTGCCATTCACACTGTGGGACAGTACAGCAAGCAAGCGC 4165
QY 210 IleGluThr----- 212
Db 4166 AGGTTACCTACAGCAGCAGCTGAGCTGGAAGAAATCGCTGTGCTCCCGGCCCA 4225
QY 213 AsnTyrThrArgIlePheLeuTyrSerGlyGluPro-----ThrTyr 226
Db 4226 GACTCAAACTGGGTGCTGCTACTTCACCCCACTCATGAGTTTGACAGCAAGCCTT 4285
QY 227 LeuGlyAsnGluThrSerValPheGlyProThrArgLysAsnLysThrLeuGlyLeuAlaIle 246
Db 4286 ATTTCAGACAGCAGCATGTGATTTATACCCCAAGAACTAGCTTTTCCAGTATTATGC 4345
QY 247 LysAspPheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuLeuSerLeu 266
Db 4346 ATATATATAGTTTACCTTATCTTTCATGAGGAGCAGCTGCTTTTGTGATTTTTTA 4405
QY 267 LeuGlnIlePhe 270
Db 4406 ATAGTCTTTT 4417

```

RESULT 3
 IS-08-290-448A-73
 Sequence 73, Application US/08290448A
 Patent No. 567654

GENERAL INFORMATION:

APPLICANT: Rogers, Bruce
 APPLICANT: Klapper, David G.
 APPLICANT: Rafnar, Thorunn
 APPLICANT: Kuo, Mei-chang
 TITLE OF INVENTION: Allergic Proteins From Ragweed and Uses
 NUMBER OF SEQUENCES: 93
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD

```

STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIORITY APPLICATION DATA:
PRIORITY APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy B. Mandragoras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-290-448A-73
Alignment Scores:
Pred. No.: 0.108 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 5.15% Indels: 93
Gaps: 16
US-10-010-050A-2_COPY_31_346 (1-316) x US-08-290-448A-73 (1-1349)
QY 35 CysPro-----ThrGlySerProIleProValMetGluLysAsp 47
Db 526 TGTCAGAGGAGCATGATTAAGTCCAGAGTGTCCACCAATTTTAAAGCAACAAAGTAT 585
QY 48 AspAspIleGluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAspLeu 67
Db 586 GGTATGCTATAATATGTTGCTGAGTTCACAAATATG----- 624
QY 68 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 87
Db 625 ATGACCATGCTGCTCAGTAAAGCTTCCAGATGGCTCGATACACCTGGCAGC 684
QY 88 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnGlyThrPheProHisLeu 107
Db 685 TCACACGTG-----ACCGTTTCCAACTGCAAAATTC----- 714
QY 108 ArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGluGly 127
Db 715 -----ACCAACCAACAAATTTGATTTATGCTGCGG 744
QY 128 IleAspAspValHisTyrPheLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
Db 745 GCTGATGACACCACTATATCAAGATTAAGGCATGCTA-----GCAACGTAAGCATTC 795
QY 148 AsnMetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGluTyrGlyIleTyr--- 166

```

Db 796 AACATGTTCCACGATACGTTGACCAAGAAAGCCTAGATGATTTGGTTTCCAA 355
 QY 167 -----TyrGluThrTyrAsnValLysAlaSerProGluLysGlyAlaGlu 181
 Db 856 GTCGTTAACACAACTACGACGATGGGAAACGTAAGCCATCGGTGAGTGGGCCCA 315
 QY 182 Thr-----TyrPheAspSerTyrAspCysSerLysPheValLeu 194
 Db 916 ACTATACTACGACCAAGGAAACAGATCTTGGCCCCCGATGATATCATCAAG----- 366
 QY 195 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnLeu 211
 Db 967 -----AAAAATGCTTACGACGACGCTGTAAGCAACCGAGTGTGATGCTGGGAAAC 1020
 QY 212 ThrAsnTyrThrArgLlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThr 231
 Db 1021 TGGAGAACAGATAGAGACTGCTTGAAATGCT----- 1053
 QY 232 SerValPheGlyProThrGly-----AsnLysThrLeuGlyLeuAlaLysArgPheTyr 250
 Db 1054 GCTATTTTCTCCATCCGGGTCTGATCCAGTGTAAACCCCTGAGCAAAAAGCAGGATG 1113
 QY 251 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuGlnLlePhe 270
 Db 1114 ATTCACGCTGAAACCA-----GAGAAAGCCGCTCTTAAGACTCACTAGTAGTGTCT 1161
 QY 271 AspAlaValLleValHisLys-----GlnPheTyrLeuPhe 282
 Db 1162 GGGTACTCTGCATGCCATCAAGAGCAGCCTGCTAGACACCTGCGCAATTCCTAAGCTTT 1221
 QY 283 TyrAsn-----PheGluTyrTyrPheLeuProMetLysPheProPheLleLysLe 299
 Db 1222 TATTAATATCATTAATACTATTTATTTTATTT-----TTGAAATTT 1263
 QY 300 ThrTyrGluGluLleProLeuProLysArgAsnLysThrLeu 313
 Db 1264 TTATATGAA-----CCATTAAGCTTCAAGTACTCTTA 1293
 RESULT 4
 US-08-290-448A-73
 ; Sequence 73, Application US/08290448A
 ; Patent No. 5698204
 ; GENERAL INFORMATION:
 ; APPLICANT: Rogers, Bruce
 ; APPLICANT: Klapper, David G.
 ; APPLICANT: Rafnar, Thorunn
 ; APPLICANT: Kuo, Mei-chang
 ; TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
 ; NUMBER OF SEQUENCES: 93
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, suite 510
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/290,448A
 ; FILING DATE: August 15, 1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/529,951
 ; FILING DATE: May 29, 1990
 ; APPLICATION NUMBER: US 07/325,365
 ; FILING DATE: March 17, 1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Amy E. Mandragouras
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: IMI-018CN

; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)227-5941
 ; INFORMATION FOR SEQ ID NO: 73:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1349 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1323
 ; US-08-290-448A-73
 Alignment Scores:
 Pred. No.: 0.108 Length: 1349
 Score: 90.50 Matches: 68
 Percent Similarity: 32.17% Conservative: 33
 Best Local Similarity: 21.66% Mismatches: 120
 Query Match: 5.15% Indels: 93
 DB: Gaps: 16
 US-10-010-050a-2_copy_31_346 (1-316) x US-08-290-448A-73 (1-1349)
 QY 35 CysPro-----ThrGlySerProLleProValMetGluGlyAsp 47
 Db 526 TGTCCAGAGGAGATGATTAATGATCCACGATGGTCCACCAATTTTAAACAAAGATGAT 585
 QY 48 AspAspLleGluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAspLeu 67
 Db 586 GGTGATGCTAATAAAGTCTGCTAGTCTTCAACAAATATG----- 624
 QY 68 LeuGlyHisLeuLysLleMetHisAspAlaLleGlyPheArgSerThrLeuThrGlyLys 87
 Db 625 ATCGACCAATGCTCGCTCAGTAAGGCTTCGATGGCTCTCGAATATCAACCTCGGACAG 684
 QY 88 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 107
 Db 685 TCACACGTC-----ACCGTTCCACCTGCAAAATTC----- 714
 QY 108 ArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaLysPhePheGluGly 127
 Db 715 -----ACCAACCAATTTGTATTATTCGCGG 744
 QY 128 IleAspAspValHisTyrLysGluAsnGlyThrLeuValGlnValAlaThrLleSerGly 147
 Db 745 GCTGATGACACCCATTAATCAAGTAAGGCATGCTA-----GCAACGGTAGCATTC 795
 QY 148 AsnMetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGluThrGlyLleTyr 166
 Db 796 AACATGTTCCACGATACGTTGACCAAGAAATGCTAGATGATTTGGTTTCCAA 855
 QY 167 -----TyrGluThrTyrAsnValLysAlaSerProGluLysGlyAlaGlu 181
 Db 856 GTCGTTAACACAACTACGACGATGGGAAACGTAAGCCATCGGTGAGTGGGCCCA 915
 QY 182 Thr-----TyrPheAspSerTyrAspCysSerLysPheValLeu 194
 Db 916 ACTATACTACGACCAAGGAAACAGATCTTGGCCCCCGATGATATCATCAAG----- 966
 QY 195 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnLeu 211
 Db 967 -----AAAAATGCTTACGACGACGCTGTAAGCAACCGAGTGTGATGCTGGGAAAC 1020
 QY 212 ThrAsnTyrThrArgLlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThr 231
 Db 1021 TGGAGAACAGATAGAGACTGCTTGAAATGCT----- 1053
 QY 232 SerValPheGlyProThrGly-----AsnLysThrLeuGlyLeuAlaLysArgPheTyr 250
 Db 1054 GCTATTTTCTCCATCCGGGTCTGATCCAGTGTAAACCCCTGAGCAAAAAGCAGGATG 1113

QY 251 TyrProPheIysProHisIeuProThrIysGluPheIeuLeuSerIeuLeuGlnIlePhe 270
 DB 1114 ATTCGAGCTGAACCA-----GGAGAACGCGTTCTTAAGACTCCTACTAGGCTT 1161
 QY 271 AspAlaValIleValHisIys-----GlnPheTyrIleuPhe 282
 DB 1162 GGTGTACTTCATGCGCATCAAGACACCTTGCTAAGACCTGCGCAATCTTAAGCTTT 1221
 QY 283 TyrAsn-----PheGluTyrTrpPheIeuProMetIysPheProPheIleIysIle 299
 DB 1222 TATTAATATCATTAATAATCTTATTATTATTATT-----TTTGATATT 1263
 QY 300 ThrTyrGluGluIleProIeuProIleArgAsnIysThrIeu 313
 DB 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293
 RESULT 5
 US-08-175-069A-73
 ; Sequence 73, Application US/08175069A
 ; Patent No. 5776761
 ; GENERAL INFORMATION:
 ; APPLICANT: Rogers, Bruce
 ; APPLICANT: Klapper, David G.
 ; APPLICANT: Ratnat, Thorunn
 ; APPLICANT: Kuo, Mei-chang
 ; TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
 ; NUMBER OF SEQUENCES: 93
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 60 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/175,069A
 ; FILING DATE: December 29, 1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/529,951
 ; FILING DATE: May 29, 1990
 ; APPLICATION NUMBER: US 07/325,365
 ; FILING DATE: March 17, 1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Amy E. Mandragouras
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: IMI-018DV
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)227-5941
 ; INFORMATION FOR SEQ ID NO: 73:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1349 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1323
 ; US-08-175-069A-73
 Alignment Scores:
 Pred. No.: 0.108 Length: 1349
 Score: 90.50 Matches: 68
 Percent Similarity: 32.17% Conservative: 33
 Best Local Similarity: 21.66% Mismatches: 120
 Query Match: 5.15% Indels: 93
 DB: 1 Gaps: 16

US-10-010-050A-2_COPY_31_346 (1-316) x US-08-175-069A-73 (1-1349)
 QY 35 CysPro-----ThrIysSerProIleProValMetGluIysAsp 47
 DB 526 TGTCCAGAGGAGATGATTAAATCAACAGATGCTCCCAATTTTAAAGACAAAGATGAT 585
 QY 48 AspAspIleGluValPheArgIeuGlnAlaProValTyrGluPheIysTyrGlyAspIeu 67
 DB 586 GGTGTATGCTATAAATGTTGCTGATGTTGATTCACAAATATG-----624
 QY 68 LeuGlyHisIeuIysIleMetHisAspAlaIleGlyPheArgSerThrIeuThrGlyIys 87
 DB 625 ATCGACCATGCTCGCTCAATGAAGCTTCGATGGAGCTGCTCGATATCAACCTCGGCGAG 684
 QY 88 AsnTyrThrMetGluTyrPyrGluIeuPheGlnIeuGlyAsnCysThrPheProHisIeu 107
 DB 685 TCACACGTG-----ACCGTTTCCAACTGCAAAATTC-----714
 QY 108 ArgProGluMetAspAlaProPheTyrCysAsnGlnIylAlaIleCysPhePheGluIyl 127
 DB 715 -----ACCCAACCAATTTGATTATTGCTCGG 744
 QY 128 IleAspAspValHisTrpIysGluAsnGlyThrIeuValGlnValAlaThrIleSerGly 147
 DB 745 GCTGATGACACCCATTATCAAGATTAAGCATGTA-----GCAACGCTAGCATTC 795
 QY 148 AsnMetPheAsnGlnMetAlaIylSerValIylGlnAspAsnGluThrGlyIleTyr 166
 DB 796 AACATGTTCAACGATCAACGATGACCAAGATGCTTAATGATGATTTGGTTTTCCTCA 855
 QY 167 -----TyrGluThrTyrAsnValIylAsnIleSerProGluIylGlyAlaIyl 181
 DB 856 GTGCTTAACAACAACTACAGACAGATGGGGAACGTACGCGATGCTGAGCTCGGCCCA 915
 QY 182 Thr-----TrpPheAspSerTyrAspCysSerIylPheValIeu 194
 DB 916 ACTATCTACAGCCAAAGGAGACAGATTCCTGGCCCCCGATGATCATCAAG-----966
 QY 195 ArgThrPheAsnIylSleuAlaGluPheGly-----AlaGluPheIylAsnIleGlu 211
 DB 967 -----AAAAATGCTTATAGCAGAGACTGTACTGTGCAACGCAAGCTCATGTCTGGAAAC 1020
 QY 212 ThrAsnTyrThrArgIlePheIeuTyrSerGlyIeuProThrTyrIleGlyAsnGluThr 231
 DB 1021 TGGAGAACAGATGACAGCTTGCTTGAATAATGCT-----1053
 QY 232 SerValPheGlyProIleGly-----AsnIylThrIeuGlyIleAlaIleIylArgPheTyr 250
 DB 1054 GCTATTTTCTCCCATCGGCTGTGATCCAGTGTCAACCCCTGAGCAAAAAGCAGGATG 1113
 QY 251 TyrProPheIysProHisIeuProThrIysGluPheIeuLeuSerIeuLeuGlnIlePhe 270
 DB 1114 ATTCAGCTGAACCA-----GGAGAACGCGTTCTTAAGACTCCTACTAGGCTT 1161
 QY 271 AspAlaValIleValHisIys-----GlnPheTyrIleuPhe 282
 DB 1162 GGTGTACTTCATGCGCATCAAGACACCTTGCTAAGACCTGCGCAATCTTAAGCTTT 1221
 QY 283 TyrAsn-----PheGluTyrTrpPheIeuProMetIysPheProPheIleIysIle 299
 DB 1222 TATTAATATCATTAATAATCTTATTATTATTATT-----TTTGATATT 1263
 QY 300 ThrTyrGluGluIleProIeuProIleArgAsnIysThrIeu 313
 DB 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293
 RESULT 6
 US-08-461-939B-73
 ; Sequence 73, Application US/08461939B
 ; Patent No. 6335019
 ; GENERAL INFORMATION:
 ; APPLICANT: Rogers, Bruce

```

APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Methods For Treating Sensitivity To A
TITLE OF INVENTION: Protein Allergen Using Peptides which include A T Cell Epitope
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,939B
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/464,000
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy B. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IWI-018CNDV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)742-4214
FAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-461-939B-73
Alignment Scores:
Pred. No.: 0.108
Score: 90.50
Percent Similarity: 32.17%
Best Local Similarity: 21.66%
Query Match: 5.15%
DB: 4
Gaps: 16
US-10-010-050a-2_COPY_31_346 (1-316) x US-08-461-939B-73 (1-1349)
QY 35 CysPro-----ThnglyserProIleProValMetGluGlyasp 47
DB 526 TGCCAGGAGGCGATGATTAGTCCACGATGGTCCCAATTTTAAAGACAAAGTAT 585
QY 48 AspAspIleGluValPheArgLeuGlnAlaProValTrpGluPheLysTrpGlyaspLeu 67
DB 586 GGTGATGCTATTAATGTCTGCTGATGTTCAAAATATGG----- 624
QY 68 LeuGluPheLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyys 87
DB 625 ATCGACCATGCTCGCTCAGTAAAGCTTCGATGGGCTGCTGATATACACCTCGGCGAGC 684
QY 88 AsnTyThrMetGluTrpTyrgluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 107

```

```

DB 685 TCACACGTG-----ACGTTTCCAACTGCAATTC----- 714
QY 108 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaIaLysPheGluGly 127
DB 715 -----ACCAACACCAATTTGATTATTTCTGGG 744
QY 128 IleAspAspValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
DB 745 GCTGATGACACCCATTTTCAAGATTAAGGCAATGCTA-----GCAACGCTAGCATTC 795
QY 148 AsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsnGlnuThnGlyIleTyr--- 166
DB 796 AACGATTTTCAACCATCAGTTCGACCAAGATGCTTGAATGATTTGGTTTTCGAA 855
QY 167 -----TyrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGlu 181
DB 856 GTGCTTAAACAACAACATCAGACAGTGGGGAACGATCGGTGGTACTCGGCCCA 915
QY 182 Thr-----TrpPheAspSerTyrAspCysSerLysPheValLeu 194
DB 916 ACTATCTACGCAAGGAAACAGATTTCTTGGCCCCGATGATATATCAATCAAG----- 966
QY 195 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 211
DB 967 -----AAAAATGCTTAGGAGAGACTGCTGATCGACACGACAGATGCTGGAAC 1020
QY 212 ThrAsnTyThrArgIlePheLeuTyrSerGlyGluProThrTyrluGlnGluThr 231
DB 1021 TGGAGAACAGATGAGACCTTGCTGAATAATGCT----- 1053
QY 232 SerValPheGlyProThrGly---AsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 250
DB 1054 GCTATTTTCTCCATCCGGGCTGTGATCCAGTCTAACCCCTGAGCAAAAAGCAGGATG 1113
QY 251 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuGlnIlePhe 270
DB 1114 ATCCGAGCTGAACCA-----GGAGAACCGCTTCTAGATCTACTAGTATGCTT 1161
QY 271 AspAlaValIleValHisLys-----GlnPheTyrLeuPhe 282
DB 1162 GGTGATCTCTCATGCTCAGGACGACCTGCTAGACACTGCGCAATTCCTAAGCTT 1221
QY 283 TyrAsn-----PheGluTyTrpPheLeuProMetLysPheProHelleLysIle 299
DB 1222 TATATATATCATTAATCTTATTTTATTTTATTT-----TTTGATATT 1263
QY 300 ThrTyrgluGluIleProLeuProIleArgAsnLysThrLeu 313
DB 1264 TTATATGAA-----CCATTACCTTCAAGTACTCTA 1293
RESULT 7
US-08-464-000-73
Sequence 73, Application US/08464000
Patent No. 6335020
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Peptides from Ragweed Pollen
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentn Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,000
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-464-000-73

Alignment Scores:
Pred. No.: 0.108      Length: 1349
Score: 90.50         Matches: 68
Percent Similarity: 32.17%      Conservative: 33
Best Local Similarity: 21.66%   Mismatches: 120
Query Match: 5.15%             Indels: 93
DB: 4                       Gaps: 16

US-10-010-050a-2_copy_31_346 (1-316) x US-08-464-000-73 (1-1349)
QY 35 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 47
DB 526 TGTCCAGGAGGCGATGATTAGTCCACGATGCTGCCATTTTAAAGCAACAAAGTGAT 585
QY 48 AapAapIleGluValPheArgLeuGlnAlaProValTTrpGluPheLysTyrGlyAspLeu 67
DB 586 GGTGATGCTATTAAGTCTGCTGTAGTACAAATATGG----- 624
QY 68 LeuGlyHisLeuLysIleMetHisAapAlaIleGlyPheArgSerThrLeuThrGlyLys 87
DB 625 ATCGACCAATGCTCGCTGAGTAAGGCTTCGATGGGCTGCTGATATACACCTGGGACGC 684
QY 88 AsnTyrThrMetGluTTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 107
DB 685 TCACACGCG-----ACCGTTTCCACCTGCMAATTC----- 714
QY 108 ArgProGluMetAapAlaProPheTTrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 127
DB 715 -----ACCAACACCAATTTGATTATTGCTGGG 744
QY 128 IleAapAapValHisTTrpLysGlnAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
DB 745 GCTGATGACACCAATATCAAGATTAAGGCAATGCTA-----GCAACGCTAGCAATTC 795
QY 148 AsnMetPheAsnGlnMetAlaLysTTrpValLysGlnAapAapGlnTThrGlyLeuTyr--- 166
DB 796 AACATGTTCCACGATCACTGACCAAAAGATGCTAGATGATTTGGGTTTTCCAA 855
QY 167 -----TyrGluThrTrpAenValLysAlaSerProGluLysGlyValGln 181
DB 856 GTGCTTAACAACAACACTACACAGATGCGGAAGCTACGCGCATGGGTGGAGCTCGACCCCA 915
QY 182 Thr-----TTrpPheAapSerTyrAapCysSerLysPheValLeu 134
DB 111 -----

```

```

DB 916 ACTATACTCAGCAAGGGAAGAGATTCTTCGCCGCCCGATGATATCATCAG----- 966
QY 195 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 211
DB 967 -----AAAAATGCTTTAGCCAGGAGCTGACTGGCAACGAGAGATGCTGGGAAC 1020
QY 212 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnLutThr 231
DB 1021 TGGAGAACAGATGAGAGCTTCGTTGAATAATGCT----- 1053
QY 232 SerValPheGlyProThrGly---AsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 250
DB 1054 GGTATTTTTCCTCCATCCGGGCTGATCCAGTGCTAACCCCTGCAGCAAAAGCAGGATG 1113
QY 251 TyrProPheLysProHisLeuProThrLysGlnPheLeuSerLeuLeuGlnIlePhe 270
DB 1114 ATTCAGCTGAACCA-----GGAGAGCCGTTCTTAAGACTCAGTAGTGCT 1161
QY 271 AapAlaValIleValHisLys-----GlnPheTyrLeuPhe 282
DB 1162 GGTGACTCTCATGCCATCAAGAGCACTTGTAAAGCACTGGCCAAATCCTAAAGCTTT 1221
QY 283 TyrAen-----PheGluTyrTrpPheLeuProMetLysPheProPheIleLysIle 299
DB 1222 TATATATATCATTAATATCTTATTTATTTATTT-----TTGATATT 1263
QY 300 ThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 313
DB 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTTA 1293

RESULT 8
US-08-290-448A-58
Sequence 58, Application US/08290448A
Patent No. 5676954
GENERAL INFORMATION:
APPLICANT: ROGERS, Bruce
APPLICANT: Klapner, David G.
APPLICANT: Rafter, Thorum
APPLICANT: Kuo, Wei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```

TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1328
US-08-290-448A-58

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	0.143	1328	68	28	113	117
Percent Similarity:	89.50					
Best Local Similarity:	29.45%					
Query Match:	20.86%					
DB:	5.09%					
	1					
	15					

US-10-010-050a-2_copy_31_346 (1-316) x US-08-290-448A-58 (1-1328)

```

QY 35 CysPro-----ThrglySerProIleProValMetGluGlyAsp 47
    |||||
DB 505 TGTCCAGGAGCATGATTAGTCCACAGCATGTCCACCAATTTTAAAGACAAAGATGAT 564
    |||||
QY 48 AspAspIleGluValPheArgLeuGlnAlaProValTPGluPheLysTYrGlyAspLeu 57
    |||||
DB 565 GGTATGCTAATAATGTTGCTGCTAGTTCACAAATATG----- 503
    |||||
QY 68 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 37
    |||||
DB 604 ATCCACCATGCTGCTGCTAGTAAGGCTTCGATGGCTGCTGATACACCTCGGAGC 563
    |||||
QY 88 AsnTYrThrMetGluTPYrTYrGluLeuPheGlnLeuGlyLysCysThrPheProHisLeu 107
    |||||
DB 664 TCACACGTG-----ACCGTTTCCAACTGCCAAATTC----- 593
    |||||
QY 108 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaIleLysCysPhePheGluGly 127
    |||||
DB 694 -----ACCCAAACCAATTTGATTATGCTCGGG 723
    |||||
QY 128 IleAspAspValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
    |||||
DB 724 GCTGATGACACCCATATCAAGATTAAGGATGCTA-----GCAAGGATGACATTC 774
    |||||
QY 148 AsnMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTYr--- 166
    |||||
DB 775 AACATGTTACCCGATCAGTTGACCAAGATGCCATGATTTGGGTTTTCCAA 834
    |||||
QY 167 -----TYrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGlu 81
    |||||
DB 835 GTGCTTAACAACAACACTACGACAGATGGGAAACGACCATCGGTGCTAGCTGCCCCCA 894
    |||||
QY 182 Thr-----TrpPheAspSerTYrAspCysSerLys----- 931
    |||||
DB 895 ACTATACTACGCAAGGAGACAGATCTTCCGCCCCGATATATCATCAAGAAATGTC 954
    |||||
QY 192 -----PheValLeuAlaGluThrPheAsn 998
    |||||
DB 955 TTAGCGAGAGCTGTTACTGGCAACGACAGTCATGTCGGAACTGGAAACAGATTA 1014
    |||||
QY 199 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGlnThrAsnTYrThrArgIlePhe 218
    |||||
DB 1015 GACTTGCTTGAATAATGCTGCT-----ATTATT 2041
    |||||
QY 219 LeuTYrSerGlyGluProThrTYrLeuGlyAsnGluThrSerValPheGlyProThrGly 238
    |||||
DB 1042 CTCGCCATCCGGGCTGTCATCCAGTCAAGTCAACCCCTGAGACA----- 1080
    |||||
QY 239 AsnLysThrLeuGlyLeuAlaIleLysArgPheTYrTYrProPheLysProHisLeuPro 258
    |||||
DB 1081 -----AAAGCAAGGATGATTTCACTGAAACCA 1107
    |||||
QY 259 ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLys--- 277
    |||||
DB 1108 --GAGAGAACCGCTTCAAGATCTCACTAGTAGTGCTGCTACTCTCATGCCATCAAGCA 1164
    |||||

```

```

QY 278 -----GlnPheTYrLeuPheTYrAsn-----PheGluTYr 287
    |||||
DB 1165 GCACCTTGCTAAGACCTGGCCGATTCCTTAAGCTTTATATATCATTAATACTTATT 1224
    |||||
QY 288 TrpPheLeuProMetLysPheProPheIleLysIleThrTYrGluGluIleProLeuPro 307
    |||||
DB 1225 TATTTTAT-----TTGATATTTTATATGAA-----CCA 1254
    |||||

RESULT 9
US-08-290-448A-58
Sequence 58 Application US/08290448A
Patent No. 5698204
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragoras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IM1-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURES:
NAME/KEY: CDS
LOCATION: 1..1328
US-08-290-448A-58

Alignment Scores:
Pred. No.: 0.143 Length: 1328
Score: 89.50 Matches: 68
Percent Similarity: 29.45% Conservative: 28
Best Local Similarity: 20.86% Mismatches: 113
Query Match: 5.09% Indels: 117
DB: 1 Gaps: 15

US-10-010-050a-2_copy_31_346 (1-316) x US-08-290-448A-58 (1-1328)
QY 35 CysPro-----ThrglySerProIleProValMetGluGlyAsp 47

```

```

Db      |||||
505  TGTCCAGAGGAGCATGATTAAAGTCCACAGATGGTCCACCAATTTTAAAGCAACAAGTAT 564
QY      |||||
48  AspAsp11egluValPheArgLeuGlnAlaProValTTPgluPheLeuTyrlAspLeu 67
Db      |||||
565  GGTGATGCTAAATATGTTGCTGTAGTCTCACAATAATG----- 603
QY      |||||
68  LeuGln11eLeuLys11eMetHisAspAla11eglyPheArgSerThrLeuThrGlyLys 87
Db      |||||
604  ATCGACCAATGCTCGCTCAGTAAAGCTTCAGATGGGCTCTCGATATACCTCGGAC 653
QY      |||||
88  AsnTyrThrMetGluTTPtyrGluLeuPheGlnLeuGlyAsnCySerThrPheProHisLeu 137
Db      |||||
664  TCACACGTG-----ACCGTTCCAACTGCAATTC----- 693
QY      |||||
108  ArgProGluMetAspAlaProPheTTPtyrCyAsnGlnGlyAlaAlaCyPhePheGlnGly 127
Db      |||||
694  -----ACCCACACCAATTTGATATATGCTCGGG 723
QY      |||||
128  IleAspAspValHisTTPtyrGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
Db      |||||
724  GGTATGACACCATTAATCAAGATTAAGCATGCTA-----GCAACGGTAGCATTC 774
QY      |||||
148  AsnMetPheAsnGlnMetAlaLysTTPvalLysGlnAspAsnGluThrGly11eTyr--- 166
Db      |||||
775  AACATGTTCCACGATCAGCTGACCAAGAAATGCTAGATGATGATTTGGGTTTCCAA 834
QY      |||||
167  -----TyrGluThrTTPAsnValLysAlaSerProGluLysGlyAlaGlu 181
Db      |||||
835  GTCGTTAAACAACACTACAGACAGATGGGAAAGTAACGCAATGCTGAGTGGCCCA 894
QY      |||||
182  Thr-----TTPheAspSerTyrAspCySerLys----- 191
Db      |||||
895  ACTATACTCAGCCAGGAGACAGATTTCTCCGCCCATGATATCATGTAAGAAAAATGTC 954
QY      |||||
192  -----PheValLeuArgThrPheAsn 198
Db      |||||
955  TTAGCGAGAGCTGTGATCTGCACAGCAGATGCTGATGCTGGAACCTGGAGAAACGATAA 1014
QY      |||||
199  LysLeuAlaGluPheGlyAlaGluPheLysAsn11egluThrAsnTyrThrArg11ePhe 218
Db      |||||
1015  GACTTGTCTTGAATATGTTGCT-----ATTTT 1041
QY      |||||
219  LeuTyrSerGlyGluPheProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 218
Db      |||||
1042  CTCCCATCCGGGTGATCCAGTGTCTAACCCCTAGCAA----- 1080
QY      |||||
239  AsnLysThrLeuGlyLeuAla11eLysArgPheTyrTyrProPheLysProHisLeuPro 218
Db      |||||
1081  -----AAAGCAGGATGATTCAGCTGAACCA----- 1107
QY      |||||
259  ThrLysGluPheLeuLeuSerLeuLeuGln11ePheAspAlaVal11eValHisLys--- 277
Db      |||||
1108  ---GGAAGAGCGTTTAAAGCTCAGTAAAGTCTGCTGATACCTCATGCGCAACAGGA 1164
QY      |||||
278  -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 267
Db      |||||
1165  GCACCTTGCTTAAGACCTCGGCAATTCCTAAGCTTTTAATAATCACTAATAATCTATTT 1224
QY      |||||
288  TTPheLeuProMetLysPheProPhe11eLys11eThrTyrGluGlu11eProLeuPro 317
Db      |||||
1225  TATTTTATTT-----TTTGAATTTTATATGAA-----CCA 1254
QY      |||||
308  IleArgAsnLysThrLeu 313
Db      |||||
1255  TTACGTTCAAGTACTCTA 1272

```

```

?  APPLICANT: Klapper, David G.
?  APPLICANT: Rafnar, Thorunn
?  APPLICANT: Kuo, Mei-chang
?  TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
?  NUMBER OF SEQUENCES: 93
?  CORRESPONDENCE ADDRESS:
?  ADDRESSEE: LAHIVE & COCKFIELD, LLP
?  STREET: 60 State Street
?  CITY: Boston
?  STATE: Massachusetts
?  COUNTRY: USA
?  ZIP: 02109-1875
?  COMPUTER READABLE FORM:
?  MEDIUM TYPE: Floppy disk
?  COMPUTER: IBM PC compatible
?  OPERATING SYSTEM: PC-DOS/MS-DOS
?  SOFTWARE: Patent Release #1.0, Version #1.25
?  CURRENT APPLICATION DATA:
?  APPLICATION NUMBER: US/08/175,069A
?  FILING DATE: December 29, 1993
?  PRIOR APPLICATION DATA:
?  APPLICATION NUMBER: US 07/529,951
?  FILING DATE: May 29, 1990
?  APPLICATION NUMBER: US 07/325,365
?  FILING DATE: March 17, 1989
?  ATTORNEY/AGENT INFORMATION:
?  NAME: Amy B. Mandragouras
?  REGISTRATION NUMBER: 36,207
?  REFERENCE/DOCKET NUMBER: IMI-018DV
?  TELECOMMUNICATION INFORMATION:
?  TELEPHONE: (617)227-7400
?  TELEFAX: (617)227-5941
?  INFORMATION FOR SEQ ID NO: 58:
?  SEQUENCE CHARACTERISTICS:
?  LENGTH: 1328 base pairs
?  TYPE: nucleic acid
?  STRANDEDNESS: single
?  TOPOLOGY: linear
?  MOLECULE TYPE: cDNA
?  FEATURE:
?  NAME/KEY: CDS
?  LOCATION: 1..1328
?  US-08-175-069A-58
?
?  Alignment Scores:
?  Pred. No.: 0.143 Length: 1328
?  Score: 89.50 Matches: 68
?  Percent Similarity: 29.45% Conservative: 28
?  Best Local Similarity: 20.86% Mismatches: 113
?  Query Match: 5.09% Indels: 117
?  Gaps: 15
?
?  US-10-010-050A-2_COPY_31_346 (1-316) x US-08-175-069A-58 (1-1328)
?
QY      |||||
35  CysPro-----ThrGlySerProIleProValMetGluGlyAsp 47
Db      |||||
505  TGTCCAGAGGAGCATGATTAAAGTCCACAGATGGTCCACCAATTTTAAAGCAACAAGTAT 564
QY      |||||
48  AspAsp11egluValPheArgLeuGlnAlaProValTTPgluPheLeuTyrlAspLeu 67
Db      |||||
565  GGTGATGCTAAATATGTTGCTGTAGTCTCACAATAATG----- 603
QY      |||||
68  LeuGln11eLeuLys11eMetHisAspAla11eglyPheArgSerThrLeuThrGlyLys 87
Db      |||||
604  ATCGACCAATGCTCGCTCAGTAAAGCTTCAGATGGGCTCTCGATATACCTCGGAC 663
QY      |||||
88  AsnTyrThrMetGluTTPtyrGluLeuPheGlnLeuGlyAsnCySerThrPheProHisLeu 107
Db      |||||
664  TCACACGTG-----ACCGTTCCAACTGCAATTC----- 693
QY      |||||
108  ArgProGluMetAspAlaProPheTTPtyrCyAsnGlnGlyAlaAlaCyPhePheGlnGly 127
Db      |||||
694  -----ACCCACACCAATTTGATATATGCTCGGG 723

```

QY 128 ILeaspapValH:stRPLySGluAsnGlyThrLeuValGluValAlaThrIleSerGly 147
Db 724 GCTATAGACCCCATTTTCAAGATTAAGCATGCTA-----GCAAGCATGACATTC 774
QY 148 AsnMetPheAsnGlnMetAlaIleTyrValLysGlnAspAsnGluThrGlyIleTyr--- 166
Db 775 AACATGTTACCCGATCAGCTTGAACCAAGAAATGCTAGATGTAGATTTGGTTTTCGAA 334
QY 167 -----TyrGluThrTyrAsnValLysAlaSerProGluLysGlyValGlu 181
Db 835 GTGCTTAACAACAACATAGACAGATGGGAACGTAACCCATCGGTAGTGGGCCCA 394
QY 182 Thr-----TyrPheAspSerTyrAspCysSerLys----- 191
Db 895 ACTATCTACAGCCAGGAAGAAAGATCTTCCGCCGATATATCATCAAGAAATGTC 354
QY 192 -----PheValLeuAlaGlyThrPheAsn 398
Db 955 TTACGAGAGACTGTACTGGCAACGAGATGATGTCGGAACCTGGAGAACAGATPAA 014
QY 199 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 218
Db 1015 GACTTGCTTGAATAATGCTGCT-----ATTTT 041
QY 219 LeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 238
Db 1042 CTCCTCATCGGGTCTGATCCAGTCCATMACCCTGAGCA----- 1080
QY 239 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHISLeuPro 258
Db 1081 -----AAAGCAGGATGATTCAGCTTAACCA 1107
QY 259 ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHISLys--- 277
Db 1108 ---GAGAAGCCCTTAAAGACTCAGTACAGTACGCTGAGTACTCTCCATCCATCAAGA 1164
QY 278 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 287
Db 1165 GCACCTTGCTTAAGACCTGCGCAATTCCTAAGCTTTTATTAATATCAATAATACCTATT 1224
QY 288 TyrPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 307
Db 1225 TATTTTATT-----TTTGATTAATTATTAAGAA-----CCA 1254
QY 308 ILeArgAsnLysThrLeu 313
Db 1255 TTACGTTCAAGTACTCTA 1272
RESULT 11
US-08-461-939B-58
Sequence 58, Application US/08461939B
Patent No. 6335019
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Wei-chang
TITLE OF INVENTION: Methods For Treating Sensitivity To A
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,939B
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/464,000
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CNDV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEO ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1328
US-08-461-939B-58
Alignment Scores:
Pred. No.: 0.143 Length: 1328
Score: 89.50 Matches: 68
Percent Similarity: 29.45% Conservative: 28
Best Local Similarity: 20.86% Mismatches: 113
Query Match: 5.09% Indels: 117
Gaps: 15
US-10-010-050a-2_copy_31_346 (1-316) x US-08-461-939B-58 (1-1328)
QY 35 CysPro-----ThrGlySerProIleProValMetGluLysAsp 47
Db 505 TGTCCAGAGAGGATGATTAAGTCCACAGATGGTCCACCAATTTTAAAGACAAAGATGAT 564
QY 48 AspAspIleGluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAspLeu 67
Db 565 GGTATGCTATTAATGTTGCTGTAGTTCACCAATATAG----- 603
QY 68 LeuGlyHISLeuLysIleMetHISAspAlaIleGlyPheArgSerThrLeuThrGlyLys 87
Db 604 ATCAACCATTTGCTGCTGCTAGTAAGGCTTCCGATGGCTCGATATACACCTCGGACGC 663
QY 88 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHISLeu 107
Db 664 TCACACGCG-----ACCGTTTCCAACTCAATTC----- 693
QY 108 ArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaLysCysPhePheGluGly 127
Db 694 -----ACCCAAACCAATTTGATTTATGCTCGGG 723
QY 128 ILeaspapValH:stRPLySGluAsnGlyThrLeuValGluValAlaThrIleSerGly 147
Db 724 GCTATAGACCCCATTTTCAAGATTAAGCATGCTA-----GCAAGCATGACATTC 774
QY 148 AsnMetPheAsnGlnMetAlaIleTyrValLysGlnAspAsnGluThrGlyIleTyr--- 166
Db 775 AACATGTTACCCGATCAGCTTGAACCAAGAAATGCTAGATGTAGATTTGGTTTTCGAA 834
QY 167 -----TyrGluThrTyrAsnValLysAlaSerProGluLysGlyValGlu 181
Db 835 GTGCTTAACAACAACATAGACAGATGGGAACGTAACCCATCGGTAGTGGGCCCA 894

```

QY 182 Thr-----TrrpheAspSerTyrAspCysSerTyr-----191
Db 895 ACTATACAGCAAGGAAAGAGATTCCTGCCCCGATATATCATCAAGAAATGTC 554
QY 192 -----PheValLeuArgThrPheAsn 198
Db 955 TTACGAGAGACTGTAAGTGGCAACGAGAGTCGATCGGAACTGAGAAAGATATA 1014
QY 199 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 218
Db 1015 GACTTCTGTAATAATGCTGCT-----ATTATTT 1041
QY 219 LeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 238
Db 1042 CTCACATCCGGGCTGATCCAGTCAAGTCAACCCCTGAGCA-----1080
QY 239 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuPro 258
Db 1081 -----AAGAGAGGAGATGATCCAGCTGAGCA-----1107
QY 259 ThrLysGluPheLeuLeuSerLeuGlnIlePheAspAlaValIleValHisLys---277
Db 1108 ---GGAGAAAGCCGTTTAAAGACTACTAGTACTGCTGCTGATCTCATGATGCAAGGA 1164
QY 278 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 287
Db 1165 GCACCTTGCTAAGCACTGGCCAAATTCCTAAGCTTTTATATATATCAATAAATCTTATTT 1224
QY 288 TrpPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 307
Db 1225 TTTATTTATTT-----TTTGATATTTTATATGAA-----CCA 1254
QY 308 IleArgAsnLysThrLeu 313
Db 1255 TTACGTTCAAGTACTCTA 1272

RESULT 12
US-08-464-000-58
/ Sequence 58, Application US/08464000
/ Patent No. 6335020
/ GENERAL INFORMATION:
/ APPLICANT: Rogers, Bruce
/ APPLICANT: Klapper, David G.
/ APPLICANT: Rafnar, Thorunn
/ APPLICANT: Kuo, Mei-chang
/ TITLE OF INVENTION: Allergenic Peptides from Ragweed Pollen
/ NUMBER OF SEQUENCES: 93
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: LAHIVE & COCKFIELD, LLP
/ STREET: 60 State Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02109-1875
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/464,000
/ FILING DATE: 05-JUN-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/290,448
/ FILING DATE: 15-AUG-1994
/ APPLICATION NUMBER: US 07/529,951
/ FILING DATE: 29-MAY-1990
/ APPLICATION NUMBER: US 07/325,365
/ FILING DATE: 17-MAR-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Amy E. Mandragouras
/ REGISTRATION NUMBER: 36,207
/ REFERENCE/DOCKET NUMBER: IMI-018CN2

```

```

/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 227-7400
/ TELEFAX: (617) 227-5941
/ INFORMATION FOR SEQ ID NO: 58:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1328 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..1328
/ US-08-464-000-58

Alignment Scores:
Pred. No.: 0.143 Length: 1328
Score: 89.50 Matches: 68
Percent Similarity: 29.45% Conservative: 28
Best Local Similarity: 20.86% Mismatches: 113
Query Match: 5.09% Indels: 117
DB: Gaps: 15

US-10-010-050A-2_COPY_31_346 (1-316) x US-08-464-000-58 (1-1328)
QY 35 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 47
Db 505 TGTCCAGAGAGCATGATTAAGTCCACAGATGCTCCACCAATTTTAAACAAAGATGAT 564
QY 48 AspAspIleGluValPheArgLeuGlnIleProValITrpGluPheLysTyrGlyAspLeu 67
Db 555 GGTATGCTATTAATGTTGCTGCTGATGTTTACCAAAATATG-----603
QY 68 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 87
Db 604 ATGCACCATTCGTCGCTCAAGTAAGGCTTCGATGGCTCGATATACCTCGGACG 663
QY 88 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 107
Db 664 TCACACGTG-----ACCGTTTCAACTGCAGAAATTC-----693
QY 108 ArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGluGly 127
Db 694 -----ACCAACCAACCAATTTGATTAATGCTCGCG 723
QY 128 IleAspAspValHisTyrLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
Db 724 GCTGATGACACCATTTATCAAGATTAAGGATGCTA-----GCACGTTAGCATTC 774
QY 148 AsnMetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGluThrGlyTyr---166
Db 775 AACATGTTCAACCGATCAAGTTCACCAAGATGCTAGATGATTTGGGTTTTCCAA 834
QY 167 -----TyrGluThrTyrAsnValLysAlaSerProGluLysGlyAlaGlu 181
Db 835 GTCGTTAACAACAATCTACGACAGATGGGAAAGTACGCGATCGGTGAGTGGGCCCA 894
QY 182 Thr-----TrrpheAspSerTyrAspCysSerTyr-----191
Db 895 ACTATACAGCAAGGAAAGAGATTCCTGCCCCGATATATCATCAAGAAATGTC 954
QY 192 -----PheValLeuArgThrPheAsn 198
Db 955 TTACGAGAGACTGTAAGTGGCAACGAGAGTCGATCGGAACTGAGAAAGATATA 1014
QY 199 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 218
Db 1015 GACTTCTGTAATAATGCTGCT-----ATTATTT 1041
QY 219 LeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 238
Db 1042 CTCACATCCGGGCTGATCCAGTCAAGTCAACCCCTGAGCA-----1080

```



```

; Patent No. 6100023
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Design Assay
; FILE REFERENCE: STDU-03484
; CURRENT APPLICATION NUMBER: US/09/157,021A
; EARLIER FILING DATE: 1998-09-18
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-157-021-1

Alignment Scores:
Pred. No.: 1.37      Length: 4440
Score: 88.50      Matches: 71
Percent Similarity: 36.39%      Conservative: 56
Best Local Similarity: 20.34%      Mismatches: 115
Query Match: 5.03%      Indels: 107
DB: 3      Gaps: 20

US-10-010-050a-2_copy_31_346 (1-316) x US-09-157-021-1 (1-4440)
QY 16 LysArgPheAspPheArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCys 35
DB 563 AAGAAAGATCATTCATTGAGCTTACCGAATCTTCTCTGTGCACACGCCAGCAGT 622
QY 36 ProthrGly-----SerProIlePro 42
DB 623 CTTTCAAGAAAAATGCTTCCCTCTCCAGAAAGAGCCTCCATTAACTCCCTTCTCT 682
QY 43 ValMetGluGlyAsp-----AspAspIleGluVal-----PheArgLeu 55
DB 683 CTTAAAGCTGCTGATTTGCTGCGCATGATGACAGGTGAGAGTACAGAAAGAGAGAG 742
QY 56 GlnAlaProValTyrGluPheLysTyrGlyAspLeuLysGlnHisLeuLysIleMetHis 75
DB 743 GAGTCAGGGGCAATGAGTTGTAAGATGATGACTTT----- 778
QY 76 AspaIalIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGlu 95
DB 779 -----GATGAGCCCATGAGATTGAAAG 832
QY 96 LeuPheGlnLeuGlyAsnCySerThrPheProHisLeuArgProGluMetAspAlaProPhe 115
DB 803 GTG-----GACCTGGAGCCT--ATGGCTGCCAAGCCT 832
QY 116 TrrCysAnGlnGlyAlaAlaCysPhePheGluGlyIleAspAspValHisTrrPheGlu 135
DB 833 TGGGCAAAAGAGAGT-----GAGCCAGCAGAGAGAGTGAACAAGAGCG 877
QY 136 Asn-----GlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln 152
DB 878 GATTTCGGAAAGGAGCGGTG-----TCTTACTTAAAGAGTTTCTCCCGAT 925
QY 153 MetAlaLysTrr-----ValIleGlnAspAnGlnThrGlyIleTyrTyrGluThrTrrAsn 171
DB 926 GTCTCTGTGGACATTCATGATCAAGAGTGTATAGCAGTTTCTGTCAGAAAGTTCA 935
QY 172 ValIleAspLysPro-----GluLysGlyAlaGluThr----- 132
DB 986 GTGATTCACAGTCACCTCCATTCGTGTAAGAGGAGATGAGAGCAAGATTCACACTT 1345
QY 183 ---TrrPheAspSerTyrAsp-----CysSerLysPheValLeuArgThrPheAsn 138
DB 1046 TATTGGTTGGATGCTTATAGAGATAGTACAAACCAAGAGGTGGTGTATTTCTGTTGGG 1105
QY 199 LysLeuAlaGluPheGlyAlaGlu-----PheLysAsnIle 210

```

```

DB 1106 AAAGTTTGATTTGAATGAGCCGAGACCATGTGACTGTGTGTGATGTAATAATATC 1165
QY 211 GlnThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGlu 230
DB 1166 GAGCGAAGC-----CTTAC-----TTCCTTCCCGTGA 1195
QY 231 ThrSerValPheGlyProThrGlyAsnLysThr-----LeuGlyLeuAlaIleLysArgPhe 249
DB 1196 ATGAAATTTGATCTTAATATCGGGGAAAGAACAGCACTCCATTTCAATGAAGATGTT 1255
QY 250 TrrTyrProPheLysProHisLeuProThrIleLysGlnPheLeuLeuSerLeuLeuGlnIle 269
DB 1256 TATGAGATTTGATGAGAAATATGCAACAATATTAATATATGAG----- 1303
QY 270 PheAspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPhe 289
DB 1304 TTCAAGCTCAAGCCAGTGAAGAAAGAAC-----TATGCTTTGAGATACCTGAT 1351
QY 290 LeuProMetLysPheProPheIleLysIleThrTyrGlu-----GluIlePro 305
DB 1352 GTTCAGAAAAATCTGAGTACTTGGAGTTAAATCTCGGCTGAATGCCACAGCTTCT 1411
QY 306 LeuProIleArgAsnLysThrLeuSer 314
DB 1412 CAAGATTGAAAGAGAAACTTTTCT 1438

RESULT 15
US-09-156-842-1
; Sequence 1, Application US/09156842A
; Patent No. 6103473
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Screening
; FILE REFERENCE: STDU-03485
; CURRENT APPLICATION NUMBER: US/09/156,842A
; EARLIER FILING DATE: 1998-09-18
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-156-842-1

Alignment Scores:
Pred. No.: 1.37      Length: 4440
Score: 88.50      Matches: 71
Percent Similarity: 36.39%      Conservative: 56
Best Local Similarity: 20.34%      Mismatches: 115
Query Match: 5.03%      Indels: 107
DB: 3      Gaps: 20

US-10-010-050a-2_copy_31_346 (1-316) x US-09-156-842-1 (1-4440)
QY 16 LysArgPheAspPheArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCys 35
DB 563 AAGAAAGATCATTCATTGAGCTTACCGAATCTTCTGTGTGCACACGCCAGCAGT 622
QY 36 ProthrGly-----SerProIlePro 42
DB 623 CTTTCAAGAAAAATGCTTCCCTGTCTCCAGAAAGAGCCTCCATTAACTCCCTGTTCT 682
QY 43 ValMetGluGlyAsp-----AspAspIleGluVal-----PheArgLeu 55
DB 683 CTTAAAGCTGCTGATTTGCTGCGCATGATGACAGGTGAGAGTACAGAAAGAGAGAG 742
QY 56 GlnAlaProValTyrGluPheLysTyrGlyAspLeuLysGlnHisLeuLysIleMetHis 75
DB 743 GAGTCAGGGGCAATGAGTTGTAAGATGATGACTTT----- 778

```

```
QY 76 Aspa1a1leglyPheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGlu 35
Db 779 -----GATGAGCCCATGGAAGTTGAAAG 302
QY 96 LeuPheGlnLeuGlyAsnCysThrPheProHISLeuArgProGluMetAspAlaProPhe 115
Db 803 GTG-----GACCTGAGCCT--ATGGCTGCCAAGGCT 332
QY 116 TrpCysAsnGlnGlyAlaAlaCysPhePheGluGlyIleAspAspValHISTrpLysGlu 135
Db 833 TGGGACAAAGAGAGT-----GAGCCAGACAGAGAGTGAACAAAGAGCG 377
QY 136 Asn-----GlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln 152
Db 878 GATTCTGGGAAAGGAGCGGTG-----TCCCTACTAGAGAGTTTCTCCCGGAT 325
QY 153 MetAlaLysTrp---ValLysGlnAspAsnGluThrGlyIleTyrTyrGluThrTyrAsn 171
Db 926 GTCCTCTGTGGGACATTGATCAGAAAGGTGATGCAAGTTTCTCAGTCAAGAAATTCAA 385
QY 172 ValLysAlaSerPro-----GluLysGlyValAlaGluThr----- 182
Db 986 GTGATTCCTCACTCACCTCCCATTTGTTAAAGGGCAGATGAGAACAGATATTCACCTT 1045
QY 183 ---TrpPheAspSerTyrAsp-----CysSerLysPheValLeuArgThrPheAsn 198
Db 1046 TATGTGTTGATGCTTATGAGATCAGTACACCAACCAAGGTGATTTCTGTTTGGG 1105
QY 199 LysLeuAlaGluPheGlyValAlaGlu-----PheLysAsnIle 210
Db 1106 AAAGTTGGATTGATCAGACCGGAGACCATGTGAGCTGTGTGCATGTGAAATAATATC 1165
QY 211 GluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGlu 230
Db 1166 GAGCGAAG-----CTTAC-----TTCCTTCCCGGTGA 1195
QY 231 ThrSerValPheGlyProHISLysAsnLysThr---LeuGlyLeuAlaIleLysArgPhe 249
Db 1196 ATGAAATTTGATTAATACGGGAAAGAAACAGAACTCCAAATTCATGAAGAGATGT 1255
QY 250 TyrTyrProPheLysProHISLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIle 269
Db 1256 TATGAGGAATTTGATGAGAAATATGCAACAAATATTAATATTAAG----- 1303
QY 270 PheAspAlaValIleValHISLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPhe 289
Db 1304 TTCAAGTCTAAGCCAGTGAAGAAC-----TATGCTTTGAGATACCTGAT 1351
QY 290 LeuProMetLysPheProPheIleLysIleThrTyrGlu-----GluIlePro 305
Db 1352 GTTCCAGAAAAATCTGAGTAACTTGAAGTTAAATACTCGGCTGAATGCCACAGCTTCT 341
QY 306 LeuProIleArgAsnLysThrLeuSer 314
Db 1412 CAGATTGAAAGGAAACCTTTTCT 1438
```

Search completed: April 25, 2004, 05:51:01
Job time : 140.894 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 22:59:01 ; Search time 583.283 Seconds
(without alignments)
2301.507 Million cell updates/sec

Title: US-10-010-050a-2_copy_31_346
Perfect score: 1758
Sequence: 1 RVSGSPSRHWPVYKRRDF.....IKITYEELPIRINKTSL 316

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/gnu_1/USPTO.spool/US10010050/runat_22042004_113203_27539/app_query.fasta_1.1372
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -IOECL=0
-IOEPRX=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsu62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10010050 @CGN 1 1 819 @runat_22042004_113203_27539 -NCPU=6 -ICPU=3
-NO_MMAP -LAKEBUTERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-DEV_TIMESOUT=120 -WARN_TIMESOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1758	100.0	1486	2	AAx02855 Human zsl
2	1747	99.4	1751	2	AAz24826 Human sec
3	1744	99.2	2120	4	AAK94829 Human ful
4	1515	86.2	1038	2	AAx02866 Human ful
5	896	51.0	697	4	AAK93424 Human de3
6	896	51.0	697	4	AAK92158 Human cdn
7	886	50.4	494	4	AAI25096 Human probe #13
8	886	50.4	494	4	ABA70786 Human foe

9	886	50.4	494	4	AAI50961	AAI50961 Probe #19
10	886	50.4	494	4	ABA37272	ABA37272 Probe #15
11	886	50.4	494	4	AAK44992	AAK44992 Human bon
12	886	50.4	494	4	AAK19040	AAK19040 Human bra
13	886	50.4	494	4	ABA44658	ABA44658 Human liv
14	886	50.4	494	6	ABSI9237	ABSI9237 Human gen
15	727	41.4	506	2	AAV88347	AAV88347 EST clone
16	684.5	38.9	474	4	AAI15879	AAI15879 Probe #58
17	684.5	38.9	474	4	ABA58185	ABA58185 Human foe
18	684.5	38.9	474	4	AAI37794	AAI37794 Probe #64
19	684.5	38.9	474	4	ABA27366	ABA27366 Probe #56
20	684.5	38.9	474	4	AAK31927	AAK31927 Human bon
21	684.5	38.9	474	4	AAK06266	AAK06266 Human bra
22	684.5	38.9	474	4	ABSI1617	ABSI1617 Human liv
23	684.5	38.9	474	6	ABSO6689	ABSO6689 Human gen
24	472	26.8	497	6	ABQ55979	ABQ55979 Human ova
25	359.5	20.4	431	6	ABL79601	ABL79601 Human ova
26	241	13.7	473	6	ABL66053	ABL66053 lung canc
27	241	13.7	473	6	ABL68165	ABL68165 kidney ca
28	121	6.9	60	6	ABN37591	ABN37591 Human spl
29	104.5	5.9	2120	4	AAI60668	AAI60668 Human cdn
30	104.5	5.9	2175	4	AAI16737	AAI16737 Human cdn
31	103	5.9	1634	3	AAI59941	AAI59941 Human sec
32	96	5.5	3580	3	AAI16697	AAI16697 Human sec
33	96	5.5	3580	6	ABK35630	ABK35630 CDNA segm
34	96	5.5	4366	4	AAH57556	AAH57556 Human bra
35	95.5	5.4	8112	4	ABL06515	ABL06515 Drosophila
36	95.5	5.4	10274	4	ABL06514	ABL06514 Drosophila
37	94.5	5.4	14070	2	AAZ09525	AAZ09525 Human Apo
38	94.5	5.4	14070	2	ACA64742	ACA64742 Human Apo
39	94	5.3	1289	6	ABX70383	ABX70383 Helicobac
40	94	5.3	3717	7	ACA34695	ACA34695 Prokaryot
41	92.5	5.3	2168	4	ABL29283	ABL29283 Drosophila
42	91.5	5.2	7680	5	AAH70286	AAH70286 DNA encod
43	91.5	5.2	7680	7	ABX08846	ABX08846 Angiogene
44	91.5	5.2	7680	7	ACC86708	ACC86708 Human vas
45	91	5.2	1649	3	AAI43098	AAI43098 Arabidops

ALIGNMENTS

RESULT 1	AAx02855	standard; DNA; 1486 BP.
ID	AAx02855	
XX	AAx02855	
AC	AAx02855	
XX	AAx02855	
DT	14-MAY-1999	(first entry)
XX	14-MAY-1999	
DE	Human zslg46 DNA.	
XX	Human zslg46 DNA.	
XX	Secreted protein; zslg46; human; chromosome 13; thyroid; disease;	
KW	hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;	
KW	Hirschprung's disease; neuronal ceroid-lipofusiosis; Wilson disease;	
KW	Reiger syndrome; immunosassy; detection; anti-idiotypic antibody;	
KW	therapy; diagnostic; ss.	
XX	Human sapiens.	
OS	Human sapiens.	
XX	Human sapiens.	
FX	Key	Location/Qualifiers
FT	CDS	47..1087
FT	FT	/tag= a
FT	FT	/product= "zslg46"
XX	XX	
PN	W0905275-A1.	
XX	W0905275-A1.	
PD	04-FEB-1999.	
XX	04-FEB-1999.	
XX	24-JUL-1998;	98WO-US015431.
FF	24-JUL-1998;	98WO-US015431.
XX	24-JUL-1997;	97US-0053613P.
PR	24-JUL-1997;	97US-0053613P.
XX	24-JUL-1997;	97US-0053613P.
PA	(ZYMO) ZYMOGENETICS INC.	

XX Sheppard PO, Gilbertson DG;
 XX WPI, 1999-142930/12.
 DR P-PSDB; AAM92967.
 XX
 PT New secreted polypeptide, zsig46, and its fragments, related fusion
 PT proteins - used for diagnosis and treatment of thyroid disorders or
 PT diseases involving genes on chromosome 13.
 XX
 PS Claim 27; Page 88-90; 101pp; English.

CC This invention describes the isolation of a novel human secreted protein,
 CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in
 CC the thyroid. This product can be used to study secretion of proteins from
 CC cells and also to treat or prevent deficient expression of zsig46, which
 CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
 CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
 CC involve genes in the same region of chromosome 13 (e.g. Hirschprung's
 CC disease, neuronal ceroid-lipofuchinosis, Wilson disease and Reiger
 CC syndrome). Antibodies and other binding proteins, are used as immunoassay
 CC reagents to detect zsig46 or cells expressing it, e.g. for assessing
 CC thyroid function to produce anti-idiotypic antibodies, for affinity
 CC purification of zsig46, to screen expression libraries, to neutralise
 CC zsig46 activity and to deliver toxins, radioisotopes etc. for
 CC therapeutic or diagnostic purposes. Agonists of the product can be used
 CC to promote growth, differentiation and proliferation of specific cell
 CC types, e.g. for treating (extra)thyroid diseases or as additive to cell
 CC cultures
 XX

Sequence 1466 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.82e-188 Length: 1486
 Score: 1758.00 Matches: 316
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-010-050a-2_COPY_31_346 (1-316) X AAX02855 (1-1486)

QY 1 ArgValSerGlyIleProSerArgGHisTrpProValProTyrIleAspGlyPhe 20
 DB 137 CGGGTCTCGGGACCTCCCTCCGCGCCACCTGCGCGGTGCTTCAAGCGCTTGAAGCTTC 196
 QY 21 ArgProIleProAspProTyrCysGlnAlaIleTyrThrPheCysProthrglySerPro 40
 DB 197 CGTCCAAACCTGATCTTATTCTCAAGCTAAGTAACTTTCTGTCAACTGGCTACCT 256
 QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgIleGlnAlaProValTyr 60
 DB 257 ATCCCGATTATGAGCGGTGATGATGACATTGAAGTTTTCATTACAGCCCAAGATGG 316
 QY 61 GluPheIleTyrGlyAspIleuGlyIleuGlyIleuGlyIleuGlyIleuGlyIleuGly 80
 DB 317 GAATTTAAATATGAGACCTCTCGGACACTTGAATAATATGATGATGATGATGATGATG 376
 QY 81 ArgSerThrLeuThrGlyIleAspIleuGlyIleuGlyIleuGlyIleuGlyIleuGly 100
 DB 377 AGAAGTACATTAACTGCGAAGAACTACAGATGGAATGATGATGATGATGATGATGATG 436
 QY 101 AsnGlyThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
 DB 437 AACTGACATTTCCCATCTCCGACCTGAATGATGATGATGATGATGATGATGATGATG 496
 QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTrpIleGlyIleuGlyIleuGly 140
 DB 497 GCTGCTGCTTTTITGAGGGAATGATGATGATGATGATGATGATGATGATGATGATG 556
 QY 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrValIleGlnAsp 160
 DB 557 CAAGTAGCAACTATATCAGGAACATGTTCAACCAAAATGCGAAAGTGGTGAACAGAC 616

QY 161 AsnGluThrGlyIleTyrTyrGluThrTrpAsnValIleAspSerProGluIleGlyAla 180
 DB 617 AATGAACACAGAAATTTATATGAGACATGGAATGTAAAGCCAGCCAGAAAGGGGCA 676
 QY 181 GluThrTrpPheAspSerTyrAspCysSerIlePheValIleuArgThrPheAsnIleu 200
 DB 677 GAGACATGCTTATGATCTCTCAAGCTGTTCCAAATTTGTATAGAGACCTTTAAACAAGTGG 736
 QY 201 AlaGluPheGlyAlaGluPheIleuGlyAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220
 DB 737 GCTGAATTTGACACAGAGTTCAAGAACATAGAACCACTATACAGAAATATTTCTTTAC 796
 QY 221 SerGlyGluProThrTyrIleuGlyIleuGluThrSerValPheGlyProThrGlyAsnIle 240
 DB 797 AGTGGAGAACCTTACTATCTGGGAAATGAAACATCTGTTTGGCCCAAGGAAACAG 856
 QY 241 ThrLeuGlyIleuAlaIleIleuAspPheTyrTyrProPheIleProHisLeuProThrIle 260
 DB 857 ACTCTTGTTAGCTCCATTAAGAAATTTTATACCTTCAACCACTTGGCAACTTAA 916
 QY 261 GluPheIleuSerIleuGlnIlePheAspAlaValIleValHisIleGlnPheTyr 280
 DB 917 GAATTTCTGTGAGTCTCTTGCAGAAATTTTGATGAGAGTATGTGCAGAAACAGTTCTAT 976
 QY 281 LeuPheTyrAspPheGluTyrTyrPheLeuProMetIlePheProPheIleIleThr 300
 DB 977 TTTCTTTATATATTTGAAATATGATGATTTTACCATGAAATCCCTTTTAAATAATACA 1036
 QY 301 TYRGlulIleProLeuProIleArgAsnIleThrLeuSerGlyIleu 316
 DB 1037 TATGAAGAAATCCCTTTACTTACTATCAGAAACAAACACTCTCTGTTT 1084
 RESULT 2
 AAZ24826
 ID AAZ24826 standard; DNA; 1751 BP.
 XX
 AC AAZ24826;
 XX
 DT 02-DEC-1999 (first entry)
 XX
 XX Human secreted protein gene 16 clone HMZAD77.
 DB
 XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumor; neurodegenerative disorder; leukemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asplenia; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
 KW osteoporosis; schizophrenia; prostate; obesity; osteoclasts; thymus;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 OS Homo sapiens.
 XX
 XX WO9947540-A1.
 PN
 XX
 PD 23-SEP-1999.
 XX
 XX 18-MAR-1999; 99WO-US005804.
 PR
 XX
 PR 19-MAR-1998; 98US-0078563P.
 PR 19-MAR-1998; 98US-0078566P.
 PR 19-MAR-1998; 98US-0078573P.
 PR 19-MAR-1998; 98US-0078574P.
 PR 19-MAR-1998; 98US-0078575P.
 PR 19-MAR-1998; 98US-0078577P.
 PR 19-MAR-1998; 98US-0078578P.
 PR 19-MAR-1998; 98US-0078579P.
 PR 19-MAR-1998; 98US-0078581P.
 PR 01-APR-1998; 98US-0080312P.
 PR 01-APR-1998; 98US-0080313P.
 PR 01-APR-1998; 98US-0080314P.
 PR 01-APR-1998; 98US-0080315P.
 XX

(HUMA-) HUMAN GENOME SCI INC.

PA (HUMA-) HUMAN GENOME SCI INC.
 KX Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
 PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;
 PI Shi Y, Moore PA;
 KX WPI: 1999-562050/47.
 DR P-PSDB; AAY41323.

PT New isolated human genes, useful for diagnosis and treatment of e.g.
 PT cancers, neurological disorders, immune diseases, inflammation or blood
 PT disorders.

35 Claim 1; Page 308; 484pp; English.

3C This sequence represents a nucleic acid molecule which encodes a secreted
 3C human protein. The gene number, and the clone it is derived from, are
 3C detailed in the descriptor line. The gene can be used to generate fusion
 3C proteins by linking to the gene to a human immunoglobulin Fc portion
 3C (e.g. AA224802) for increasing the stability of the fused protein as
 3C compared to the human protein only. The invention relates to 95 novel
 3C genes and their fragments (nucleic acid sequences: AA224811-224907), amino
 3C acid sequences AA224808-224904) which are useful for preventing, treating
 3C or ameliorating medical conditions e.g. by protein or gene therapy. Also,
 3C pathological conditions can be diagnosed by determining the amount of the
 3C new polypeptides in a sample or by determining the presence of mutations
 3C in the new polynucleotides. Specific uses are described for each of the
 3C 95 polynucleotides, based on which tissues they are most highly expressed
 3C in (see AA224811 for described uses)

3C Sequence 1751 BP; 477 A; 355 C; 380 G; 523 T; 0 U; 16 Other;

3C Alignment Scores:

3C Pred. No.: 4.03e-187 Length: 1751
 3C Percent Similarity: 1747.00 Matches: 314
 3C Percent Similarity: 99.37% Conservative: 0
 3C Percent Similarity: 99.37% Mismatches: 2
 3C Query Match: 99.37% Indels: 0
 3C Gaps: 0

3C JS-10-010-050a-2_COPY_31_346 (1-316) x AA224826 (1-1751)

3C 1 ArgValSerGlyIleProSerArgHisTrpProValProCysArgPheAspPhe 20
 3C 139 CGGGTTCGGGCACTCCCTCCGGGCGCAGTGGCCGCTCAAGCGCTTGACTTC 198
 3C 21 ArgProLysProAspProCysGlyAlaIleCysProThrGlySerPro 40
 3C 199 CGTCCAAACCTGATCCTTATGCTCAAGCTAAGTATCTTCTTCCACGCTCACT 258
 3C 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrp 60
 3C 259 ATCCAGATTATGGAGGGTGAATGATGACATGAACTTTTCATTAACAAGCCCGATGAG 318
 3C 61 GluPheLysTrpGlyAspLeuLeuGlyHisIleuLysIleMetHisAspAlaIleGlyPhe 80
 3C 319 GAATTTAAATATGAGACCTCTCGGACACTTGAATTAATGATGATGACATTGAGTTC 378
 3C 81 ArgSerThrLeuThrGlyLysAsnTrpMetGlyTrpGluLeuPheGlnLeuGly 100
 3C 379 AGAAGTACATTAACTGCGCAAGAACTACACAAATGAGAAATGCGCACTTGGC 438
 3C 101 AsnCysThrPheProHisIleuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
 3C 439 AACGTACATTTCCTCCATCTCCGACCTGAATGATGATGATGATGATGATGATGATGAT 498
 3C 121 AlaAlaCysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyTrpLeuVal 140
 3C 499 GCTGCCCTCTTTTGAAGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 558
 3C 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleCysTrpValIleGlnAsp 160
 3C 559 CAAGTACCACTATATCAGAAACATGTTCAACCAATGCGCAAGTGGTGAACAGGAC 618

QY 161 AsnGluThrGlyIleTrpTrpGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180
 Db 619 AATGAAACAGAAATTATATGAGACATGATGAAAGCAGCCAGAAAGGGGCGA 678
 QY 181 GluThrTrpPheAspSerTrpAspCysSerLysPheValIleuArgThrPheAsnLysLeu 200
 Db 679 GAGCATGCTGTTGATCTCTACAGACTGTTCCAAATTTGTTAAGACCTTTTAAACAAGTTG 738
 QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTrpThrArgIlePheLeuTrp 220
 Db 739 CTTGATTTTGAAGACAGCTTCAAGACATGAGAACCACTATACAGAAATTTCTTAC 798
 QY 221 SerGlyLysProThrTrpLysGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
 Db 799 AGTGAGAACTTACTTATCTCGGAAATGAAACATCTGTTTGGGCCCAACGAAACAG 858
 QY 241 ThrLeuGlyLeuAlaIleLysArgPheTrpTrpPheLysProHisLeuProThrLys 260
 Db 859 ACTCTTGATTTAGCCATTAAGATTTTATTAACCCCTTCAAAACCAATTTGCCAACTAAA 918
 QY 261 GluPheLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGluPheTrp 280
 Db 919 GAATTTCTGTTGAGTCTCTGCAATTTTGATGAGATGATGACAAACAGTTCTAT 978
 QY 281 LeuPheTrpAsnPheGluTrpTrpPheLysProMetLysPheProPheIleLysIleThr 300
 Db 979 TTGTTTATATATTTTGAATATGCTTTTACCTATGAAATCCCTTTATTAATAATACA 1038
 QY 301 TyrGluGluIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 316
 Db 1039 TATGAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGTTTA 1086

RESULT 3
 AAK94829
 ID AAK94829 standard; cDNA: 2120 BP.
 XX AAK94829;
 AC
 XX 06-NOV-2001 (first entry)
 DT
 XX
 DE Human full-length cDNA, SEQ ID NO: 3977.
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 OS Homo sapiens.
 XX
 PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI: 2001-524255/58.
 DR P-PSDB; AAM93870.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 XX
 PS Claim 8; SEQ ID NO 3977; 1380bp + Sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have

CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a full length human cDNA of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in CD-ROM format directly
 CC from EPO

XX Sequence 2120 BP; 590 A; 409 C; 460 G; 661 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.16e-186	Length:	2120
Score:	1744.00	Matches:	314
Percent Similarity:	99.37%	Conservative:	0
Best Local Similarity:	99.37%	Mismatches:	2
Query Match:	99.20%	Indels:	0
		Gaps:	0

US-10-010-050a-2_COPY_31_346 (1-316) x AAK94829 (1-2120)

```

QY 1 ArgValSerGlyLeuProSerArgArgHisTrpProValProTyrLeuAspPhe 20
Db 142 CGGGTCTCGGSCATCCCTCCGCGCCACATGCGCGGTGCTGCAAGCGCTTGACTTC 201
QY 21 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40
Db 202 CGTCCAAACCTGATCTTATGTCAAGGTAAAGTAACTTCTGTCCTCAATGGCTCACT 261
QY 41 IleProValMetGlnGluLysAspAspAlaIleGluValPheArgLeuGlnAlaProValTrp 60
Db 262 ATCCCGAGTATGAGAGGTATGATGACCTGAACTGAACTTTTCGATTCACAGCCCGATG 321
QY 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe 60
Db 322 GATTTAAATATGAGAACCTCCCGGACACTTGAAATTATGATATGACATTGGATTTC 381
QY 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly 100
Db 382 AGAAGTACATTAACCTGCAAGAACTACACATGATGATGATGATCTTTCCAACTTGGC 441
QY 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
Db 442 AACTGTACATTTCCCATCTCCGACCTGAATGATGAGCCCTTTCGTGATATCAAGCC 501
QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTrpLysGlnAsnGlyThrLeuVal 140
Db 502 GCTGCTGCTTTTGTAGGGAATTGATGATGATGATGATGATGATGATGATGATGATG 561
QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsp 160
Db 562 CAAAGTACAACTATATCAGGAAACATGTTCAACCAATGCGCAAGTGGGTGAACAAGCAG 621
QY 161 AsnGlnThrGlyLysLeuTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180
Db 622 AATGAAACAGGAATTTATATGACATGAAAGTAAAGCCAGCCCAAAAAAGGGGCGCA 681
QY 181 GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200
Db 682 GAAACATGGTTGATCTTCCAGCCTGTTCCAAATTTGTGTAAAGACCTTTAAACAGTTTG 741
QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220
Db 742 GCGAATTTGGAGCAAGTTCAGAACATAGAAACCAATCAATACAGATATTATTTCTTAC 801
QY 221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
Db 802 AGTGGAAACCTATCTATCTGGAAATGAAACATCTGTTTGGGCAACAGAAACAAG 861
QY 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLys 260
Db 862 ACTCTTGTTGAGCCATTAAGATTTTATTAACCCCTTCAACACATTTGGCACTAA 921

```

```

QY 261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyr 280
Db 922 GATTTCTCTTACGCTCTCTGAAATTTTATGACGATGATGTCACAAACAGTTCTAT 981
QY 281 LeuPheTyrAsnPheGlnTyrTrpPheLeuProMetLysPheProPheIleLysIleThr 300
Db 982 TTGTTTATTAATTTTGAATATATGTTTAACTATGAAATTCCTTTTATTAATTAACA 1041
QY 301 TyrGluGlnIleProLeuProIleArgAsnLysTrpLeuSerGlyLeu 316
Db 1042 TATGAAAGAAATCCCTTATCTATACAGAAACAAACACTCTCGTTT 1089

```

```

RESULT 4
AAK02866
ID AAK02866 standard; DNA; 1038 BP.
AC AAK02866;
XX
XX 14-MAY-1999 (first entry)
DT
DE Human degenerate zsig46 DNA.
KW Secreted protein; zsig46; human; chromosome 13; thyroid; disease;
KW hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;
KW Hirschsprung's disease; neuronal ceroid-lipofuchnosis; Wilson disease;
KW Reiger syndrome; immunoassay; detection; anti-idiotypic antibody;
KW therapy; diagnostic; ss.
XX
OS Homo sapiens.
XX
XX WO9905275-A1.
XX
XX 04-FEB-1999.
XX
XX 24-UTL-1998; 98WO-US015431.
XX
XX 24-UTL-1997; 97US-0053613P.
XX
XX (ZYMO ) ZYMOGENETICS INC.
XX
XX Sheppard PO, Gilbertson DG;
XX
XX WPI; 1999-142930/12.
XX
XX New secreted polypeptide, zsig46, and its fragments, related fusion
XX proteins - used for diagnosis and treatment of thyroid disorders or
XX diseases involving genes on chromosome 13.
XX
XX Claim 31; Page 94-95; 101pp; English.
XX
XX This invention describes the isolation of a novel human secreted protein,
XX zsig46 encoded by a gene on chromosome 13 which is mainly expressed in
XX the thyroid. This product can be used to study secretion of proteins from
XX cells and also to treat or prevent deficient expression of zsig46, which
XX may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
XX disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
XX involve genes in the same region of chromosome 13 (e.g. Hirschsprung's
XX disease, neuronal ceroid-lipofuchnosis, Wilson disease and Reiger
XX syndrome). Antibodies and other binding proteins, are used as immunoassay
XX reagents to detect zsig46 or cells expressing it, e.g. for assessing
XX thyroid function to produce anti-idiotypic antibodies, for affinity
XX purification of zsig46, to screen expression libraries, to neutralise
XX zsig46 activity, and to deliver toxins, radioisotopes etc. for
XX therapeutic or diagnostic purposes. Agonists of the product can be used
XX to promote growth, differentiation and proliferation of specific cell
XX types, e.g. for treating (extra)thyroid diseases or as additive to cell
XX cultures
XX
XX Sequence 1038 BP; 189 A; 106 C; 179 G; 166 T; 0 U; 398 Other;
SQ
Alignment Scores:
Pred. No.: 3.3e-161 Length: 1038

```

Score: 1515.00 Matches: 267
 Percent Similarity: 85.03% Conservative: 0
 Best Local Similarity: 85.03% Mismatches: 47
 Query Match: 86.18% Indels: 0
 DB: 2 Gaps: 0

US-10-010-050a-2_COPY_31_346 (1-316) x AAK02866 (1-1036)

```

QY 2 ValSerGlyIleProSerArgArgHisTrpProValProTyrIysArgPheAspPheArg 21
DB 94 GTNMSGNGNATHCCNMWSMGMNGMNCAYTGCCNGTNCNTYAAAMGNTTGAAYTTWGN 153
QY 22 ProLysProAspProTyrCysGlnAlaIleTyrThrPheCysProThrGlySerProIle 41
DB 154 CCMARCCNGAYCCNTATATGTCARCCNAAATTAACNTTITGTCACACGWSMCCNATH 213
QY 42 ProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrpGlu 61
DB 214 CCNGTATAGARGGAGGAYGAYATHGARGTNTTGMNTNCARGCCNGTWTGGAR 273
QY 62 PheLysTyrGlyAspLeuLeuGlyHisIleuLysIleMetHisAspAlaIleGlyPheArg 81
DB 274 TTYAARTAYGNGAYTNTYNGNCAYTNAARATGATGAYGAYGCAATGGATTTWGN 313
QY 82 SerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsn 101
DB 334 MSNACNTYNAACNGNAAATAATTAACNATGARGTGTGARGTNTTTCARITWGNARAY 393
QY 102 CysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAla 121
DB 394 TGYACNTTCCNCAAYTMTGNCNGARATGAGGNCCTTGTGTGAYAACARGGNGCN 453
QY 122 AlaCysPhePheGluGlyIleAspAspValHisTrpIlyGluLeuValGln 141
DB 454 GCNTGYTITTYTGARGGATGAGAYGTCNAYTGARGARAAAYGGACVNTGTCTCAR 523
QY 142 ValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleSTPValIlyGlnAspAsn 161
DB 514 GTNCCNACNATHMSGNMAATATGTTTAAATCAATGGCNAARTGGGTTNARCAARGATAY 573
QY 162 GluThrGlyIleTyrTyrGluThrTrpAsnValIlyAsaSerProGluLysGlyAlaGlu 181
DB 574 GARACNGNATHHTTYYTAYGARACNTGGAAYGTNAARGCWNMNCNGARAAARGGNGCAR 633
QY 182 ThrTrpPheAspSerTyrAspCysSerIlyPheValLeuArgThrPheAsnIlyLeuAla 201
DB 634 ACNTGGTITGAYMSNTAYGATGTGWSNAAARTTGTNTNMGNACNTTAAAYAAAYTNGCN 693
QY 202 GluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSer 221
DB 694 GARTTGGGNGCGARITTYAARAAYATHGARGCNAATTAACMGNATHHTTYYTNTATWSN 753
QY 222 GlyIleuProThrTyrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIlySer 241
DB 754 GGNARCCNACNTYTYTNGGAAAYGARAACWSNNTTGGCCNACNGGAAAYAAARACN 813
QY 242 LeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisIleuProThrIlyGlu 261
DB 814 YTNAGNTTNGNATHAARMGNTTATATATACNTTAAACCCNAYTNCNACNAAARGAR 873
QY 262 PheLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeu 281
DB 874 TTYTNTVNTMNSYNTNTNCARAYHTTGTGAGCNGTNTHTGNCAYARCARITTYATAYTN 933
QY 282 PheTyrAsnPheGluTyrTrpPheLeuProMetLysPheProPheIleLysIleThrTyr 301
DB 934 TTYTAAAYATTTGARTYTGTTTNTNCNATGAAARTTCCNTTATTAARATATACNTAY 953
QY 302 GluGluIleProLeuProIleArgAsnLysThrLeuSerGly 315
DB 994 GARCARATHCCNTYTNCCNATHMGNAAAYAAARACNTTWSNGSN 1035

```

```

AAK93424
ID AAK93424 standard; cDNA; 697 BP.
XX AC
XX AAK93424;
XX AC
XX 06-NOV-2001 (first entry)
XX DE
XX Human cDNA clone representative sequence, SEQ ID NO: 1884.
XX KW
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX OS
XX Homo sapiens.
XX PN
XX EPI130094-A2.
XX PD
XX 05-SEP-2001.
XX PF
XX 07-JUL-2000; 2000EP-00114089.
XX PR
XX 08-JUL-1999; 99JP-00194486.
XX PR
XX 11-JAN-2000; 2000JP-00118774.
XX PR
XX 02-MAY-2000; 2000JP-00183765.
XX PA
XX (HELI-) HELIX RES INST.
XX PI
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX PI
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX DR
XX WPI; 2001-524255/58.
XX XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
XX PT
XX in genetic manipulation.
XX PS
XX Example 11; SEQ ID NO 1884; 1380bp + Sequence Listing; English.
CC CC
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesized by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence was used as the representative sequence
CC from a human clone which was used in homology searches to identify the
CC clone. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in CD-ROM format directly from
CC EPO
XX SQ
XX Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;

```

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:
1	697	1,969.91	896.00
Percent Similarity:		92.51%	Mismatches: 3
Best Local Similarity:		90.91%	Indels: 8
Query Match:		50.97%	Gaps: 2

US-10-010-050a-2_COPY_31_346 (1-316) x AAK93424 (1-697)

```

QY 1 ArgValSerGlyIleProSerArgArgHisTrpProValProTyrIysArgPheAspPhe 20
DB 142 CGGAGTCTGGGCAATCCCTCCGCGCACTGCGCGGTGCTGCAAGCGCTTGACTTC 201
QY 21 ArgProLysProAspProTyrCysGlnAlaIleTyrThrPheCysProThrGlySerPro 40
DB 202 CGTCCAAAACCTGATCTTATTTGTCAAGCTAAATATACCTTCTGTCCAACTGGCTCACT 261
QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrp 60
DB 262 ATCCCAATTATGAGGAGGTGATGATGACACTGAAGTATTTTCGATTACAAAGCCAGATATG 321

```

QY	61	GlutPheIysTyrGlyAspPheIeuGlyHisIeuYsIleMetHisaspAlaIleGlyPhe	80
Db	322	GAATTTAAATATGGAACACTCCTGGGACACTTGAAATTAATGATGATGCCATTGGATTC	381
QY	81	ArgSerThrIeuThrGlyIysAsnTyrThrMetGluTyrPylGluIeuPheGlnIeuGly	100
Db	382	AGAAAGACATTAATCTGGCAAGAACTACACAAATGGAATGGTATGAACTTTTCCAACTTGGC	441
QY	101	AsnCySThrPheProHisIeuIleuArgProGluMetAspAlaProPheTyrCyAsnGlnGly	120
Db	442	AACGTGACATTTCCCATCTCCGACCTGMAATGATGATGCCCTTTCTGGTGAATCAAGGC	501
QY	121	AlaAlaCySPhelheGluGlyIleAspAspValHisTyrPylGluAsnGlyThrIeuVal	140
Db	502	GCTGCCTGCTTTTTSAGAGGAATGATGATGTTCACTGGAAAGAAAATGGGACATTAGT	561
QY	140	IleGlnValAlaIleThrIleSer-GlyAsnMetPheAsnGlnMetAla-IleTyrPylValysGln	159
Db	562	TCAAGTAGCAACTATTCAGGAAAACATGTTCAACCAATATGGCAAAAGTGGGTGAACAA	621
QY	160	--AspAsnGluThrGlyIleTyrTyrGlu-ThrTyrAsnValIysAlaSerProGluTyr	178
Db	622	GGACATNTGGAACAAAGGAAATTTATATGAGAACATGGAATTTTAAAGCCANCCCAANNA	681
QY	178	S---GlyAlaGlu 181	
Db	682	AAANGGGGCAAAA 694	
RESULT 6			
AAK92158			
XX	ID	AAK92158 standard, cDNA, 697 BP.	
XX	AC	AAK92158;	
XX	DT	06-NOV-2001 (first entry)	
XX	DE	Human cDNA 5'-end sequence, SEQ ID NO: 618.	
XX	XX	Human, full length cDNA; cDNA synthesis; oligo-capping; ss.	
XX	OS	Homo sapiens.	
XX	XX	EP1130094-A2.	
XX	XX	05-SEP-2001.	
XX	XX	07-JUL-2000; 2000EP-00114089.	
XX	XX	08-JUL-1999; 99QP-00194486.	
XX	XX	11-JAN-2000; 2000DP-00118774.	
XX	XX	02-MAY-2000; 2000DP-00185765.	
XX	XX	(HELI-) HELIX RES INST.	
XX	XX	Ota T, Nishikawa T, Isogai T, Hayaishi K, Iehi S, Kawai Y;	
XX	XX	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX	XX	WPI; 2001-524255/58.	
XX	XX	830 Primers useful for synthesizing full length cDNA clones and their use	
XX	XX	in genetic manipulation.	
XX	XX	Claim 2; SEQ ID NO 618; 1380bp + Sequence listing; English.	

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is the nucleotide sequence of the 5'-end of

	CC	a cDNA provided in the invention Note: The sequence data for this patent
	CC	did not form part of the printed specification, but was obtained in CD-
	CC	ROM format directly from EPO
XX	SQ	Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;
	Alignment Scores:	
	Pred. No.:	1_9e-91 Length: 697
	Score:	896.00 Matches: 170
	Percent Similarity:	92.51% Conservative: 3
	Best Local Similarity:	90.91% Mismatches: 8
	Query Match:	50.97% Indels: 6
	DB:	Gaps: 2
	US-10-010-050A-2_COPY_31_346 (1-316) X AA932158 (1-697)	
QY	1	ArgValSerGlyIleProSerArgArgHisTrpProValProTyrIlySarGheAspPhe 20
Db	142	CGGGTCTCGGGATCCTCCCTCCGGGCACACAGCGCGGCCCTTGCAAGCGCTTGACTTC 201
QY	21	ArgProIyvePRioAspProTyrCYneGlnAlaystrThrPheCySPromngIysePro 40
Db	202	CGTCCAAACCTGATCCTATTATTTGTAACCTAAAGTATACCTTTCTGTCACAACCTGGCTCACT 261
QY	41	IleProValMetGluGIYAspAspAsprIleGluValPheArgLeuGlnAlaprovalTrp 60
Db	262	ATCCGAGTATATGAGGGGAGATGATGACACTGAAGTTTTGCATTACAAAGCCCCAGTATGG 321
QY	61	GluPheLyseTyrGIYAspLeuleuGIYHisteuylsIleMethIsaPaIalleGIYPhe 80
Db	322	GAATTTAAATAATGAGAACCCTCCTGGGACACTTGAAAANTATAGCVATGATGCCATTGGATTC 381
QY	81	ArgSerThrLeuThrgIYLYsaantYrThMeTGluTrpIYrGluLeupeGInLeuGIY 100
Db	382	AGAAGTACTTAACTTAGCGCAAGACTACACATAGGATGGATGATGAACCTTTCOAACCTGGC 441
QY	101	AsnCYSThPheProHisIeuArpProGluMetLspAlapropHeTrPYCaSengInGIY 120
Db	442	AACCTGACATTTCCCACATCCGACCTGAATGATGGATGCCCTTCTGTGTAATCAAGGC 501
QY	121	AlaAlaCYphePheGIUGIYIleAspAspValHisTrpLySGluAenGIYThrLeu-Va 140
Db	502	GCIGCTCGCTTTTITTTGAGGGAATTCATATGTTCACTCGAAGGAATAATGGACATTAACT 561
QY	140	IgInValAlatnrIleSer-GlyAsnMetPheaSngImetAla-LysTrpValLySGln 159
Db	562	TCAAGTAGACAATATATCAAGGAACAAGTTCACCAAAATGGCAAAGTGGTGAAACAA 621
QY	160	--AspaSngIuThrgIYIleTyrTrpGlu-ThrTpAsnVallySaIaseerProGluIy 178
Db	622	GGAACAATNGAACAAAGAAATTTATATAGAACAATGGAATTTTAAAGCCANCCCAANAA 681
QY	178	s---GIYAlaGln 181
Db	682	AAANGGGCACAAA 694
	RESULT 7	
	AA125096	
ID	AA125096	standard; DNA; 494 BP.
AC	AA125096;	
XX	XX	
DT	12-OCT-2001	(first entry)
DE	Probe #15029	for gene expression analysis in human cervical cell sample.
XX	XX	
KW	Probe; human; microarray; gene expression; cervical epithelial cell;	
XX	cervical cancer; ss.	
OS	Homo sapiens.	
XX	XX	
PN	WO200157278-A2.	

09-AUG-2001.
30-JAN-2001; 2001WO-US000670.
04-FEB-2000; 2000US-0180312P.
26-MAY-2000; 2000US-0207456P.
30-JUN-2000; 2000US-00608408.
03-AUG-2000; 2000US-00632366.
21-SEP-2000; 2000US-0234687P.
27-SEP-2000; 2000US-0236359P.
04-OCT-2000; 2000GB-00024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488901/53.
Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human cervical epithelial cells.
Claim 25; SEQ ID NO 15029; 487bp; English.
The present invention relates to human single exon nucleic acid probes (SENP). The present sequence is one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.57e-90 Length: 494
Score: 886.00 Matches: 163
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.40% Indels: 0
Gaps: 0
US-10-010-050a-2_COPY_31_346 (1-316) x AAI25096 (1-494)
QY 154 AlalysrtrpvallysglnaspaengluThrGlyIleTyTrgltuThrTPAsnVallys 173
DB 3 GCAAAAGTGGGTGAAACAGACATGAAACAGAAATTATATGACATGGAATGTAAAA 62
QY 174 AlaserProGluLyGlyAlaGluThrTPheAspSerTyAspCySerLyPheVal 193
DB 63 GCCAGCCCAAGAAAGGGGCGAGACATGCTTATCTCCGACTGCTCCAAATTTGTG 122
QY 194 LeuarGrThrPheAsnLyLeuAlaGluPheGlyAlaGluPheLyAsnIleGluThAsn 213
DB 123 TTTAAGACCTTAAACAGCTTGGCTGAATTTGAGAGAGTTCAAGAACATGAAACCAAC 182
QY 214 TyrThrArgIlePheLeuTySerGlyGluProThrTyLeuGlyAsnGluThSerVal 233
DB 183 TATACAGAAATATTTCTTACAGTGAACCTACTTATCTGGGAATGAAACATCTGTT 242
QY 234 PheGlyProThrGlyAsnLyThLeuGlyLeuAlaIleLyAsnArgPheTyTyProPhe 253
DB 243 TTTGGGCGCAACAGAAACAGACTCTGTTTACCAATAAAAATTTTATTAACCCCTTC 302
QY 254 LysProHisLeuProThrLyGluPheLeuSerLeuGlnIlePheAspAlaVal 273
DB 303 AAACCAATTGGCACTAAAGAAATTCGTGAGCTCTTGCAAAATTTTGATGCAAGTG 362
QY 274 IleValHisLyGlnPheTyLeuPheTyAsnPheGluTyTrpPheLeuProMetLys 293
DB 363 ATGTGCAAAACAGTCTATTTGTTTATATAATTTTGAATATTTGTTTAAACATGAAA 422

QY 294 PheProHelleLyGlieThrTyTrgltuIleProLeuProIleArgAsnLyThLeu 313
DB 423 TTCCCTTTATTAATAATACATATGAAAGAAATCCCTTACCTATCAGAAACAAACACTGC 482
QY 314 SerGlyLeu 316
DB 483 TCTGGTTTA 491
RESULT 8
ABA70786
ID ABA70786 standard; DNA; 494 BP.
AC ABA70786;
DT 01-FEB-2002 (first entry)
DE Human foetal liver single exon nucleic acid probe #19091.
XX Human, foetal liver; gene expression; single exon nucleic acid probe; ss.
XX Homo sapiens.
XX WO200157277-A2.
PN 30-JAN-2001; 2001WO-US000669.
PD 09-AUG-2001.
PF 30-JAN-2001; 2001WO-US000669.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-483447/52.
Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver.
Claim 4; SEQ ID NO 19091; 639bp + Sequence Listing; English.
The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a single exon nucleic acid probe of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.57e-90 Length: 494
Score: 886.00 Matches: 163
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.40% Indels: 0
Gaps: 0
US-10-010-050a-2_COPY_31_346 (1-316) x ABA70786 (1-494)
QY 154 AlalysrtrpvallysglnaspaengluThrGlyIleTyTrgltuThrTPAsnVallys 173
DB 3 GCAAAAGTGGGTGAAACAGACATGAAACAGAAATTATATGACATGGAATGTAAAA 62

QY 174 AAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSerIysPheVal 193
 DB 63 GCCAGCCCGAGAAAGGGGGCAGAGACATGGTTGATTCTTACGACTGTTCCAAATTGTG 122
 QY 194 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn 213
 DB 123 TTAAAGACCTTTTAAACAAGTTGGCTGAATTTGGAGCAGAGTTCAAGAACATTAAGAAC 182
 QY 214 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 233
 DB 183 TATACAGAAATATTTCTTTACAGTGAAGAACCTACTATCTGGAAATGAAACATCTGTT 242
 QY 234 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPhe 253
 DB 243 TTGGGGCCCAACAGGAACAAAGACTCTGGTTAGCCATAAAAAGATTATTAACCCCTTC 302
 QY 254 LysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGluIlePheAspAlaVal 273
 DB 303 AAACCAACATTTGCCAATTAAGAAATTTCTGTGAAGTCTTGCAAATTTTATGACAGTG 362
 QY 274 IleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeuProMetLys 293
 DB 363 ATTGTGCACAAACAGTCTATTTGTTTATATTTGAATTTGTTTACCTATGAAG 422
 QY 294 PheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 313
 DB 423 TTCCCTTTATTAATAATACATATGAAGAAATCCCTTACTATCAGAAACAAACACTC 482
 QY 314 SerGlyLeu 316
 DB 483 TCTGGTTTA 491

RESULT 9

AA150961
 ID AA150961 standard; DNA; 494 BP.

AC AA150961;

DT 17-OCT-2001 (first entry)

DE Probe #19647 used to measure gene expression in human placenta sample.

KM Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

OS Homo sapiens.

PN MO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-0063236P.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX PT gene expression in human placenta.

XX PS Claim 25; SEQ ID NO 19647; 654bp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENPs).

CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders

XX SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1-57e-90	Length:	494
Score:	886.00	Matches:	163
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	50.40%	Indels:	0
DB:	4	Gaps:	0

US-10-010-050a-2_COPY_31_346 (1-316) x AA150961 (1-494)

QY 154 AAlaYstTPValIlyGlnAspAsnGluThrGlyIleTyrTyrGluThrTrpAsnValLys 173
 DB 3 GCAAAAGTGGTGAACACAGACATGAACAGGAATTTATTAAGACATGGAATGA 62
 QY 174 AAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSerIysPheVal 193
 DB 63 GCCAGCCCGAGAAAGGGGGCAGAGACATGGTTGATTCTTACGACTGTTCCAAATTGTG 122
 QY 194 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn 213
 DB 123 TTAAAGACCTTTTAAACAAGTTGGCTGAATTTGGAGCAGAGTTCAAGAAACATGA 182
 QY 214 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 233
 DB 183 TATACAGAAATATTTCTTACAGTGAAGAACCTTACTTGTGGGAAATGAACATCTGTT 242
 QY 234 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPhe 253
 DB 243 TTGGGGCCCAACAGGAACAAAGACTCTGGTTAGCTTGCACAAATTTTATTAACCCCTTC 302
 QY 254 LysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGluIlePheAspAlaVal 273
 DB 303 AAACCAACATTTGCCAATTAAGAAATTTCTGTGAAGTCTTGCAAATTTTATGACAGTG 362
 QY 274 IleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeuProMetLys 293
 DB 363 ATTGTGCACAAACAGTCTATTTGTTTATATTTGAATTTGTTTACCTATGAAG 422
 QY 294 PheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 313
 DB 423 TTCCCTTTATTAATAATACATATGAAGAAATCCCTTACTATCAGAAACAAACACTC 482
 QY 314 SerGlyLeu 316
 DB 483 TCTGGTTTA 491

RESULT 10

ABA37272
 ID ABA37272 standard; DNA; 494 BP.

AC ABA37272;

DT 23-JAN-2002 (first entry)

DE Probe #15738 for gene expression analysis in human heart cell sample.

KM Human; gene expression; heart; microarray; vascular system; probe;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

OS Homo sapiens.

PN MO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000666.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 XX
 DR Single exon nucleic acid probes for analyzing gene expression in human
 XX hearts.
 PT
 PS Claim 4; SEQ ID NO 15738; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pcr_sequences
 XX
 SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.57e-90 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 50.40% Indels: 0
 DB: Gaps: 0
 US-10-010-050A-2_COPY_31_346 (1-316) x ABA37272 (1-494)
 QY 154 AAlaLysTrpValIlysgLInspaengluInrGlyIleTyTrpGluThrTrpAsnValIys 173
 Db 3 GCMAAGTGGGTGMAACAGACATGAACAGAAATTTATTATGACATGATGTAATAA 62
 QY 174 AAlaSerProgluIlysgIlylaGluThrTrpPheAspSerTyTrpAspCysSerIysPheVal 193
 Db 63 GCCAGCCCAAGAAAGGGGGCAGACATGGTTGATTCCTACACATGTTCCAAATTTGIG 122
 QY 194 ILeuArgThrPheAsnIlyseuIagluPheGlyIlylaGluPheIysAsnIleGIuThrAsn 213
 Db 123 TTAAGGACCTTTAAACAAGTTGGCTGAATTTGGAGCAGAGTTCAAGAAACATAGAACCAAC 182
 QY 214 TyrThrArgIlePheLeuTyTrpSerGlyGluProThrTyTrpLeuGlyAsnGluThrSerVal 233
 Db 183 TATACAAAGATATTTCTTACAGTGAAGAACCTACTATCTGGAAATGAAACATCTGTT 242
 QY 234 PheGlyProThrGlyAsnIlyseuThrLeuGlyLeuValIleIysAsnIleGlyPheTyTrpPhe 253
 Db 243 TTGGGGCCACAGAAACAGAACTCTTGGTTTGCCATTAAGAAAGATTTATTAACCCCTTC 302
 QY 254 IysPheIleuProThrIlysgIlyPheLeuSerLeuLeuGluIlePheAspAlaVal 273
 Db 303 AAACCACTTGGCACTAAAGAAATTCGTTGAGTCTCTGCAATTTTGTATGACAG 362
 QY 274 ILeuAlaIlyseuIlnPheTyTrpLeuPheTyTrpAsnPheGluTyTrpPheLeuProMetIys 293
 Db 363 ATTGCGACAAACAGTTCTATTTGTTTATTAATTTGAATATTGTTTACTATGAAA 422

QY 294 PheProPheIleIysIleThrTyTrpGluGluIleProLeuProIleArgAsnIlyThrLeu 313
 Db 423 TTCCCTTTATTAAATACATATGAAAGAAATCCCTTACTATCGAAGAAACAAACACTC 482
 QY 314 SerGlyLeu 316
 Db 483 TCTGGTTTA 491
 RESULT 11
 AAK44992
 ID AAK44992 standard; DNA; 494 BP.
 XX
 AC AAK44992;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 19549.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000666.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488900/53.
 XX
 DR Single exon nucleic acid probes for analyzing
 XX gene expression in human bone marrow.
 PT
 PS Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human bone marrow.
 XX
 PS Example 4; SEQ ID NO 19549; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention
 XX
 SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.57e-90 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 50.40% Indels: 0
 DB: Gaps: 0
 US-10-010-050A-2_COPY_31_346 (1-316) x AAK44992 (1-494)
 QY 154 AAlaLysTrpValIlysgLInspaengluInrGlyIleTyTrpGluThrTrpAsnValIys 173
 Db 3 GCMAAGTGGGTGMAACAGACATGAACAGAAATTTATTATGACATGATGTAATAA 62
 QY 174 AAlaSerProgluIlysgIlylaGluThrTrpPheAspSerTyTrpAspCysSerIysPheVal 193

Db 63 GCCAGCCGAGAAAAGGGGAGAGACAGGTTTGAATTCCTACGACTGTTCCAAATTGTG 122
 QY 134 LeuArgThrPheAsnIleuAlaGluPheGlyAlaGluPheIleuAsnIleGluThrAsn 213
 Db 123 TTAAGGACCTTTAAACAAGTGGCTGAATTTGAGAGAGGTTCAAGAACATGAACCAAC 182
 QY 214 TyrThrArgIlePheIleuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 233
 Db 183 TATACAGAAATTTCTTTACAGTGAAGACCTACTTATCTGGGAAATGAACATCTGTT 242
 QY 234 PheGlyProThrGlyAsnIleuThrLeuGlyLeuAlaIleuAsnIleuPheTyrTyrProPhe 253
 Db 243 TTTGGGCCACAGAGAAACAGACTCTGTGTTAGCAATAAAGATTTATATCCCTTC 302
 QY 254 LysProHisIleuProThrIleuGluPheLeuSerIleuGluIlePheAsnIleuVal 273
 Db 303 AAACCAATTTGCCAACAATTAAGAAATTCGTGAGCTCTTGCATAATTTGATGACAGTG 362
 QY 274 ILeuAlaHisIleuGlnPheTyrLeuPheTyrAsnIleuTyrTyrPheLeuProMetLys 293
 Db 363 ATTGGCAACAACAGTCTATTTGTTTATATTTTGAATATTTGTTTATCCATGAAA 422
 QY 294 PheProPheIleuIleuIleuThrTyrGluGluIleProLeuProIleuArgAsnIleuThrLeu 313
 Db 423 TTCCCTTTTATTAATAATACATATGAGAAATCCCTTAACATGAGAAACAAACACTC 482
 QY 314 SerGlyLeu 316
 Db 483 TCTGTTTA 491

RESULT 12

AAK19040
ID AAK19040 standard; DNA; 494 BP.

AC AAK19040;
 DT 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 19031.

DE Human; brain expressed exon; gene expression analysis; probe; microarray;
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
 ss.

XX Homo sapiens.

XX WO200157275-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001MO-US000667.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-0063236P.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human
 XX brains.
 XX PS Example 4; SEQ ID NO 19031; 650bp + Sequence Listing; English.
 XX CC The present invention provides a number of single exon nucleic acid
 XX probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention

XX SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,57e-90 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 50.40% Indels: 0
 DB: 4 Gaps: 0

US-10-010-050a-2_copy_31_346 (1-316) x AAK19040 (1-494)

QY 154 AlaIleuTyrValIleuGlnAsnIleuThrGlyIleTyrTyrGluThrTyrAsnValLys 173
 Db 3 GCAAGTGGGTGAGAAACAGGACATGAACAGAAATTTATATGACATGAAATGTAAA 62
 QY 174 AlaSerProGluIleuGlyAlaGluThrTyrPheAspSerTyrAspCysSerIleuPheVal 193
 Db 63 GCCAGCCGAGAAAAGGGGAGAGACAGGAGATGTTGATTCCTACGACTGTTCCAAATTGTG 122
 QY 194 LeuArgThrPheAsnIleuAlaGluPheGlyAlaGluPheIleuAsnIleGluThrAsn 213
 Db 123 TTAAGGACCTTTAAACAATTTGCTGGAATTTGAGAGAGTTCAAGAACATGAACCAAC 182
 QY 214 TyrThrArgIlePheIleuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 233
 Db 183 TATACAGAAATTTCTTTACAGTGAAGACCTACTTATCTGGGAAATGAACATCTGTT 242
 QY 234 PheGlyProThrGlyAsnIleuThrLeuGlyLeuAlaIleuAsnIleuPheTyrTyrProPhe 253
 Db 243 TTTGGGCCACAGAGAAACAGACTCTGTGTTAGCAATAAAGATTTATATCCCTTC 302
 QY 254 LysProHisIleuProThrIleuGluPheLeuSerIleuGluIlePheAsnIleuVal 273
 Db 303 AAACCAATTTGCCAACAATTAAGAAATTCGTGAGCTCTTGCATAATTTGATGACAGTG 362
 QY 274 ILeuAlaHisIleuGlnPheTyrLeuPheTyrAsnIleuTyrTyrPheLeuProMetLys 293
 Db 363 ATTGGCAACAACAGTCTATTTGTTTATATTTTGAATATTTGTTTATCCATGAAA 422
 QY 294 PheProPheIleuIleuIleuThrTyrGluGluIleProLeuProIleuArgAsnIleuThrLeu 313
 Db 423 TTCCCTTTTATTAATAATACATATGAGAAATCCCTTAACATGAGAAACAAACACTC 482
 QY 314 SerGlyLeu 316
 Db 483 TCTGTTTA 491

RESULT 13
 ID ABS44658 standard; DNA; 494 BP.
 AC ABS44658;
 DT 25-FEB-2003 (first entry)

XX Human liver single exon probe, SEQ ID NO 19648.
 XX DE Human; single exon nucleic acid probe; liver; cirrhosis;
 XX KW hyperlipoproteinemia; hyperlipidaemia; hypercholesterolaemia;
 XX KW coronary heart disease; ss.
 XX OS Homo sapiens.
 XX PS WO200157273-A2.
 XX PD 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US000664.
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488898/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human adult liver.
 PS Claim 4; SEQ ID NO 19648; 658bp; English.
 CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
 CC measuring human gene expression in a sample derived from human adult
 CC liver, comprising one of 13109 defined nucleotide sequences given in the
 CC specification (or complements/fragments). The probe hybridizes at high
 CC stringency to a nucleic acid molecule expressed in the human adult liver.
 CC (I) may be used for predicting, measuring and displaying gene expression
 CC in samples derived from human adult liver. The genes identified may be
 CC involved in genetic liver diseases such as cirrhosis,
 CC hyperlipoproteinemia, hyperlipidaemia and hypercholesterolaemia which is
 CC associated with coronary heart disease. ABS25011-ABS51005 represent human
 CC liver single exon nucleic acid probes of the invention. Note: The
 CC sequence information for this patent does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.57e-90 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 50.40% Gaps: 0
 DB: 4
 US-10-010-050A-2_COPY_31_346 (1-316) x ABS44658 (1-494)
 CY 154 AAlaySTpVallLySGlnAspAangLuthrGlylletYrYrGluThrTPAsnVallys 173
 DB 3 GCAAGTGGGTGAACAGAGACATGAAACAGAAATTATATGAGACATGGAAATGTAATA 62
 CY 174 AAlaserProGluYrGlyAlaGluThrTrpPheAspSerTYrAspCYsSerLysPheVal 193
 DB 63 GCCAGCCCAAGAAAAGGGGGCAGACATGTTGATTCCTACGATCTGTTCCAAATTTGTC 122
 CY 194 LeuAryrThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnLysGluThrAsn 213
 DB 123 TTAAGGACCTTTAACAAGTTGGTGAATTGGACAGAGATTCAAGAACATAGAAACCAAC 182
 CY 214 TyrThrArgGliePheLeuTyrSerGlyGlnProThrTyrLeuGlyAsnGluThrSerVal 233
 DB 183 TATACAAATAATTTCTTTACAGTGAACCTTACTTATCTGGGAAATGAAACATCTGTT 242
 CY 234 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysAlaGlyPheTyrTrpPhe 253
 DB 243 TTTGGGCCAACAAGAAACAAGACTCTTGTTAGCCARAAAAAATTTATACCCCTTC 302
 CY 254 LysProHisLeuProThrLysGluPheLeuLysSerLeuLysGluIlePheAspAlaVal 273
 DB 303 AAACCACTTGGCCAACTAAAGAAATTTCTGTGAGTCTCTTCAAAATTTTGATGACGTG 362

CY 274 lIeValHlEluYsGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeuProMetLys 293
 DB 363 ATTGTGCACAAAGAGTCTATTTGTTTATTTGAAATTTGATTTTATCCATGAAA 422
 CY 294 PheProPheLleLysIleThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 313
 DB 423 TTCCCTTTATTAATAATACATATGAAGAAATCCCTTTACTATCGAAACAAACGCTC 482
 CY 314 SerGlyLeu 316
 DB 483 TCTGCTTTA 491
 RESULT 14
 ABS19237
 ID ABS19237 standard; DNA; 494 BP.
 XX
 AC ABS19237;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe ORF from lung SEQ ID NO 19228.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; IID;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW HermanSky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenier syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001MO-US000665.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 XX
 DR Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples.
 PT
 PS Claim 4; SEQ ID NO 19228; 634bp; English.
 XX
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC ; the novel set of probes which hybridise at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample
 CC collected from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably

CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, CC having a fragment identical to the predicted exon, the probe is included CC in the above mentioned microarray; assigning exons to a single gene, CC comprising (a) identifying exons from genomic sequence by the method CC above and (b) measuring the expression of each of the exons in several CC tissues and/or cell types using hybridisation to a single exon CC microarrays having a probe with the exon, where a common pattern of CC expression of the exons in the tissues and/or cell types indicates that CC the exons should be assigned to a single gene; a peptide comprising one CC of 12011 sequences, mentioned in the specification, or encoded by the CC probe/open reading frames (ORF). The probes are used for gene expression CC analysis, and for identifying exons in a gene, particularly using human CC lung derived mRNA and for the study of lung diseases such as asthma, lung CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky- CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, CC Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The CC present sequence is a single exon probe open reading frame of the CC invention. Note: The sequence data for this patent did not form part of CC the printed specification, but was obtained in electronic format directly CC from WIPO at ftp.wipo.int/pub/published_pat_sequences

XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.57e-90	Length:	494
Score:	886.00	Matches:	163
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	50.40%	Indels:	0
DB:	6	Gaps:	0

US-10-010-050A-2_COPY_31_346 (1-316) x ABS19237 (1-494)

QY 154 AAlaYsTrpValYlYsGlnAspAsnGluThrGly11eYrYrGluThrTrpAsnValYs :.73
Db 3 GCAAGGCGGTGAACAGCAATGAAACAGAAATTTATTATGACATGAAATGTPAAA 62
QY 174 AAlaSerProGluYsGlyAlaGluThrTrpPheAspSerYrAspCySerYsPheVal :.93
Db 63 GCCAGCCCAAGAAAGGGGGGAGAGACATGTTGATTCCTACGACTCTTCCAATTTGTG :.22
QY 194 LeuAArgThrPheAsnYsLeuAlaGluPheGlyAlaGluPheYsAsn1eGluThrAsn :.23
Db 123 TTAAAGACCTTTAAACAAGTTGGCTGAATTTGGAGCAAGATTCAAGACATAGAAACCAAC :.82
QY 214 TyrThrArg11ePheLeuYrSerGlyGluProThrYrLeuGlyAsnGluThrSerVal :.233
Db 183 TATCAAGAAATATTTCTTTACAGTGAAGAACTTACTTATCTGGGAAATGAAACATCTGTT :.242
QY 234 PheGlyProThrGlyAsnYsThrLeuGlyLeuAla1eYsAspPheYrYrProPhe :.253
Db 243 TTTGGGCAACAGCAAGAAACAAGCTCTGTGTTACCAATAAAACATTTTATVACCCCTTC :.302
QY 254 YsPheProHsLeuProThrYsGluPheLeuSerLeuGln1ePheAspAlaVal :.273
Db 303 AAACCAACATTTGGCAACTAAGAAATTTCTGTGAGCTCTTGCACAAATTTTGAAGCAGTG :.362
QY 274 11eValHs1eYsGlnPheYrLeuPheYrAsnPheGluYrTrpPheLeuProMetYs :.293
Db 363 ATTGTGCAACAAGCTCTTATTTGTTTAAATTTTGAATTTGGTTTATCCATATGAAA 422
QY 294 PheProHelleYs11eThrYrGluGlu1eProLeuPro11eAsnYsThrLeu :.313
Db 423 TTCCTTTTATTAATAACATATGAAGAATCCCTTACTATCAAGAAACAAACACATC 482
QY 314 SerGlyLeu 316
Db 483 TCTGCTTAA 491

RESULT 15
AAV88347
ID AAV88347 standard; cDNA; 506 BP.

XX AAV88347;

DT 12-FEB-1999 (first entry)

DE EST clone GB814.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemokinesis; chemokinesis; haemostasis; gene therapy; thrombolytic;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN MO9845437-A2.

PD 15-OCT-1998.

PF 10-APR-1998; 98WO-US006956.

PR 10-APR-1997; 97US-00837312.

XX (GENEY) GENETICS INST INC.

PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;

PI Spaulding V, Agostino MJ;

DR WFI; 1999-070078/06.

PT New polynucleotides encoding human secreted proteins - derived from e.g.
PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,
PT pituitary, retina and colon cDNA libraries.

XX Claim 1, Page 366; 641pp; English.

XX The present sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene therapy
XX

SQ Sequence 506 BP; 171 A; 86 C; 119 G; 130 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.56e-72	Length:	506
Score:	727.00	Matches:	133
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	41.35%	Indels:	0
DB:	2	Gaps:	0

US-10-010-050A-2_COPY_31_346 (1-316) x AAV88347 (1-506)

QY 121 AAlaCyPhePheGluGly11eAspAsnValHs1eYrYsGluThrVal :.140
Db 99 GCTGCTGCTTTTGAAGAAATTTGATGATGTTCACTGGAAGAAATGGAATTAGTT 158
QY 141 GlnValAlaThr11eSerGlyAsnMetPheAsnGlnMetAlaYsTrpValYsGlnAsp :.160
Db 159 CAAAGTGAACATATATCAGGAAACATGTTCAACCAATGGAAGTGGTGAACAGGAC 218
QY 161 AsnGluThrGly11eYrYrGluThrTrpAsnValYsAlaSerProGluYsGlyAla :.180

```
Db      219 AATGAAACAGGAAATTTATTATGAGACATGAAATGTAATAAGCCAGGCCAGAAAAGGGGCA 278
Qy      181 GIuThrTPpheAspSerTyrAspCysSerIysPheValIeuArgThrPheAsnIysIeu 200
Db      279 GAGACATGCTTGGATTCTTACGACTGTTCCAAATTTGGTTAAGGACCTTTAACAGTTG 318
Qy      201 AIAgiUpheGIyAlaGIuPheIysAsnIleGIuThrAsnTyrThrArgIlePheIeuTyr 220
Db      339 GCTGAATTTGGAGCAGAGTTCAAGAACATAGAAACCACTATACAGAAATATTCTTTAC 398
Qy      221 SerGIyGIuProThrTyrLeuGIyAsnGIuThrSerValPheGIyProThrGIyAsnIys 240
Db      399 AGTGAAGAACTTACTTATCTGGGAATGAAACATCTGTTTGGGCCAACAGAAACAG 458
Qy      241 ThrLeuGIyLeuAlaIleIysArgPheTyrTyrProPhe 253
Db      459 ACTCTTGTTAGCCATAAAAGATTTATTACCCCTTC 497
```

Search completed: April 24, 2004, 23:26:28
Job time : 589.283 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 23:02:32 ; Search time 380.792 Seconds

(without alignments)
2352.639 Million cell updates/sec

Title: US-10-010-050A-2_COPY_1_30
Perfect score: 169
Sequence: 1 MRRGAGARGRASMCWMLALMLAVPGMS 30

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODBI=frame+ p2n.model -DEV=xlh
-O=/cg2.1/USPTO.spool/US10010050/runat.22042004.113204.27556/app.query.fasta_1.1372
-DB=EST -OFMT=fasta -SUFFIX=first -MINMATCH=0.1 -IOOPT=0 -IOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biom62 -TRANS=human0.cdi -LIST=45
-LOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10010050 @CGN 1 1 6283 @runat.22042004.113204.27556 -MCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-BEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_esthum:*
5: em_esthum:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_p1n:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_ror:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	169	100.0	565	9	AUT280382
2	169	100.0	607	12	BM849262
3	169	100.0	623	13	BQ417448
4	169	100.0	676	12	BG820113
5	169	100.0	787	12	B1223533
6	169	100.0	847	13	BQ719074
7	169	100.0	869	12	B1090566
8	169	100.0	923	13	BX331615
9	169	100.0	1098	10	BF982158
10	169	100.0	1201	9	AL546472
11	169	100.0	1222	13	BQ897986
12	161	95.3	924	14	CA488543
13	156	92.3	1077	13	BX342662
14	142	84.0	589	10	BF977311
15	135	79.9	555	14	CA397574
16	119	70.4	273	9	AA300650
17	73	43.2	383	14	CB808908
18	73	43.2	424	14	CB795968
19	73	43.2	427	14	CB548129
20	73	43.2	1338	13	BU792679
21	72.5	42.9	127	28	CG378302
22	72.5	42.9	497	29	CG375783
23	72.5	42.9	587	10	BG077284
24	72.5	42.9	713	28	CG30780
25	72.5	42.9	730	29	CG108116
26	72.5	42.9	732	29	CG108118
27	72.5	42.9	825	29	CG453851
28	72.5	42.9	830	29	CG230524
29	72.5	42.9	834	29	CG230534
30	72.5	42.9	860	29	CG610458
31	72.5	42.9	2940	11	AK028566
32	72	42.6	651	29	AG105984
33	72	42.6	742	14	CF876196
34	72	42.6	855	14	CB904966
35	72	42.6	1355	10	BE880806
36	71.5	42.3	1025	10	BE567958
37	71	42.0	569	29	CE827309
38	71	42.0	632	13	BX844586
39	71	42.0	641	14	CB215456
40	70.5	41.7	840	28	AQ898634
41	70.5	41.7	401	9	AA108220
42	70.5	41.7	427	9	AA062254
43	70.5	41.7	547	12	BG101391
44	70.5	41.7	547	12	BG101262
45	70.5	41.7	591	14	CF106239

ALIGNMENTS

RESULT 1
AUT280382
LOCUS AUT280382 NIESE2 Homo sapiens cDNA clone NIESE2001763 5', mRNA
DEFINITION
SEQUENCE
ACCESSION AUT280382
VERSION AUT280382.1 GI:28299609
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 565)

Db 52 ATGGCGCGGGCGGGCGGGCTCGGGAGCGCGCTTCTGCTGCTGGGCCCTGCGCTG 111

Qy 21 LeuTripleuAlaValAlaProGlyTTrpSer 30

Db 112 CTTTGCTCGCGGTGCTTCGGGGCTGTGCC 141

RESULT 8

LOCUS BX331615

DEFINITION BX331615 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens CDNA clone CS0DB009YC01 5-PRIME, mRNA sequence.

ACCESSION BX331615

VERSION BX331615

KEYWORDS EST.

ORGANISM Homo sapiens (human)

REFERENCE Li, W.B., Gruber, C., Jessee, J. and Polayes, D. Full-length cDNA libraries and normalization Unpublished (2001)

JOURNAL

COMMENT BP 191 91006 EVRY cedex - France
Genoscope - Centre National de Sequencage
Email: secrete@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7238.f For more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DB009AB01Q1&cluster=7238.f. Contact : Feng Liang Email : fliang@lifetech.com URL : Corporation 1600
http://fulllength.invitrogen.com/Invitrogen 1600
Paraday Avenue Genoscope sequence ID : CS0DB009AB01Q1.

FEATURES

source

1..923

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DB009YC01"

/tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"

/clone_id="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"

/note="First strand cDNA was primed with a NotI-oligo (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 2,06e-07 Length: 923

Score: 169.00 Matches: 30

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 13 Gaps: 0

US-10-010-050a-2_copy_1_30 (1-30) x BX331615 (1-923)

Qy 1 MetAArgArgGlyAlaGlyAlaAlaArgGlyAArgAlaSerTPCySTrPalAlaLeuAlaLeu 20

Db 118 ATGGCGCGGGCGGGCGGGCTCGGGAGCGCGCTTCTGCTGCTGGGCCCTGCGCTG 177

Qy 21 LeuTripleuAlaValAlaProGlyTTrpSer 30

Db 178 CTTTGCTCGCGGTGCTTCGGGGCTGTGCC 207

RESULT 9

LOCUS BF982158

DEFINITION BF982158 1098 bp mRNA linear EST 23-JAN-2001

ACCESSION BF982158

BF982158

VERSION BF982158.1 GI:12384970

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1098)

AUTHORS NIH-MGC http://mgs.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@imail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.jnl.gov
plate: LLM10105 row: 9 column: 19
High quality sequence stop: 693.

FEATURES

source

1..1098

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4400298"

/tissue_type="duodenal adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_id="NIH_MGC_88"

/note="Organ: small intestine; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 2,59e-07 Length: 1098

Score: 169.00 Matches: 30

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 10 Gaps: 0

US-10-010-050a-2_copy_1_30 (1-30) x BF982158 (1-1098)

Qy 1 MetAArgArgGlyAlaGlyAlaAlaArgGlyAArgAlaSerTPCySTrPalAlaLeuAlaLeu 20

Db 35 ATGGCGCGGGCGGGCGGGCTCGGGAGCGCGCTTCTGCTGCTGGGCCCTGCGCTG 94

Qy 21 LeuTripleuAlaValAlaProGlyTTrpSer 30

Db 95 CTTTGCTCGCGGTGCTTCGGGGCTGTGCC 124

RESULT 10

LOCUS AL546472

DEFINITION AL546472 1201 bp mRNA linear EST 31-MAY-2003

ACCESSION AL546472

VERSION AL546472

KEYWORDS EST.

ORGANISM Homo sapiens (human)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1201)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D. Full-length cDNA libraries and normalization Unpublished (2001)

JOURNAL On Feb 15, 2001 this sequence version replaced gi:12879620.

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seque@genoscope.cns.fr Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODI030CE010P1aCluster=7238.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CSODI030CE010P1.

FEATURES

source

```

1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI030Y001"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dt)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

Alignment Scores:
Pred. No.: 2,91e-07 Length: 1201
Score: 169.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-010-050a-2_COPY_1_30 (1-30) x A1546472 (1-1201)

QY 1 MetAgaAGGlyAlaGlyAlaAlaAargGlyAargAlaSerTPCyETripAlaLeuAlaLeu 20
Db 63 ATGCGCGGCGGCGCGCGCGCGCGCGCGCTCGGAGACGCGCTCTGCTGCTGCGCGCTG 122

QY 21 LeuTripleuAlaValProGlyTyrSer 30
Db 123 CTTTGCTCTCGCGGTTCGCGGCTGCTGTC 152

RESULT 11

LOCUS BQ897986 1222 bp mRNA linear EST 16-AUG-2002

DEFINITION AGENCOURT 8482480 lupski_dorsal_root ganglion Homo sapiens cDNA

clone IMAGE:6185883 5', mRNA sequence.

ACCESSION BQ897986

VERSION BQ897986

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NIH-MGC http://mgs.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-riemail.nih.gov

Tissue Procurement: Dr. James R. Lupski

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.lnl.gov

Plate: L1AM13577 row: 9 column: 04

High quality sequence start: 7

High quality sequence stop: 277.

Location/Qualifiers

1..1222

FEATURES

source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6185883"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="MDH10B"
/clone_lib="Lupski_dorsal root ganglion"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGCTCG-3' and
5'-GACTAGTCTGATCGGAGCGCCCTT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

```

ORIGIN

Alignment Scores:
Pred. No.: 2,97e-07 Length: 1222
Score: 169.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-10-010-050a-2_COPY_1_30 (1-30) x BQ897986 (1-1222)

QY 1 MetAgaAGGlyAlaGlyAlaAlaAargGlyAargAlaSerTPCyETripAlaLeuAlaLeu 20
Db 17 ATGCGCGGCGGCGGCGGCGGCGGCGGCGCTCGGAGACGCGCTCTGCTGCTGCGCGCTG 76

QY 21 LeuTripleuAlaValProGlyTyrSer 30
Db 77 CTTTGCTCTCGCGGTTCGCGGCTGCTGTC 106

RESULT 12

LOCUS CA488543

DEFINITION AGENCOURT 10808884 MAPL Homo sapiens cDNA clone IMAGE:6720309 5',

mRNA sequence.

ACCESSION CA488543

VERSION CA488543.1 GI:24950702

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NIH-MGC http://mgs.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-riemail.nih.gov

Tissue Procurement: Kristi A. Egland, Ira Pastan

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.lnl.gov

Plate: L1AM14279 row: 3 column: 21

High quality sequence stop: 637.

Location/Qualifiers

1..924

FEATURES

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6720309"

/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,

hTERT-HME1, INCap"
/lab host="EMDH10B"
/clone lib="MAPCL"
/note="Vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kritici A. Eglund, James J. Vincent, Robert Strusberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:

pred. No.:	1.36e-06	Length:	924
Score:	161.00	Matches:	29
Percent Similarity:	96.67%	Conservative:	0
Best Local Similarity:	96.67%	Mismatches:	1
Query Match:	95.27%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050A-2_COPY_1_30 (1-30) x CA488543 (1-924)

1 MetAragGlyAlaGlyAlaAlaArgGlyAArgAlaSerTyrCysTyrPalaLeuAlaLeu 20
|||||
16 ATGTGGCGGGGGCGGGCGGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 75

21 LeuTyrPleuAlaValAlaProGlyTyrSer 30
|||||
76 CTTTGCTCGCGGGGTGCTCGGGCTGTCTC 105

RESULT 13
EX342662

DEFINITION BX342662 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
EX342662
BX342662 Homo sapiens cDNA clone CS0DL004YP03 5-PRIME, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1077)
I.J.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f. For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DL004CH02Q0P1&cluster=7238.f. Contact :
Peng Liang Email : filiang@lifetech.com URL :
http://fulllength.invitrogen.com/invitrogen Corporation 1600
Paradey Avenue Genoscope sequence ID : CS0DL004CH02Q0P1.

FEATURES
SOURCE

1. 1077
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL004YP03"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
distilled with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

pred. No.:	5.39e-06	Length:	1077
Score:	156.00	Matches:	30
Percent Similarity:	96.77%	Conservative:	0
Best Local Similarity:	96.77%	Mismatches:	0
Query Match:	92.31%	Indels:	1
DB:	13	Gaps:	0

US-10-010-050A-2_COPY_1_30 (1-30) x BX342662 (1-1077)

1 MetAragGlyAlaGlyAlaAlaArgGlyAArgAlaSerTyrCysTyrPalaLeuAlaLeu 20
|||||
115 ATGCGGGGGGGCGGGCGGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174

20 uLeuTyrPleuAlaValAlaProGlyTyrSer 30
|||||
175 GCTTTGCTCGCGGGGTGCTCGGGCTGTCTC 205

RESULT 14
BF977311

LOCUS BF977311 589 bp mRNA linear EST 22-JAN-2001
DEFINITION 602146485F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4305501 5',
mRNA sequence.
BF977311

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 589)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@emil.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCMI172 row: 1 column: 22
High quality sequence stop: 586.

FEATURES
SOURCE

1. 589
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4305501"
/tissue_type="melanotic melanoma, high MDR"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_62"
/note="Organ: Skin; Vector: pDNR-LIB (Clontech); Site 1:
Stit (ggcgccctggcc); Site 2: Stit (ggccatattggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGAGCGGCGGCGGACATG-dT(30)NN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

ORIGIN

Alignment Scores:

pred. No.:	6.62e-05	Length:	589
Score:	142.00	Matches:	29
Percent Similarity:	96.67%	Conservative:	0

Best Local Similarity: 96.67%
 Query Match: 84.02%
 DB: 10
 Gaps: 0

US-10-010-050a-2_COPY_1_30 (1-30) x BF977311 (1-555)

QY 1 MetAArgAGlYAlaGlyAlaAlaArgAlaSerTrpCySerTrpAlaLeuAlaLeu 20
 Db 161 ATCGCGCGGGCGCGCGCGGT-CGCGGAGCGCGCTTCTGCTGCTGCGCGCGCTG 219
 QY 21 LeuTrpLeuAlaValAlaProGlyTrpSer 30
 Db 220 CTTTGCTCGCGGTGCTCGGCGCTGCTC 249

RESULT 15
 CA397574

LOCUS
 DEFINITION CA397574 555 bp mRNA linear EST 06-NOV-2002
 cs92h10.y2 Human Retinal pigment epithelium/choroid cDNA
 (Un-normalized, unamplified): cs Homo sapiens CDNA clone cs92h10
 5', mRNA sequence.

ACCESSION CA397574
 VERSION CA397574
 KEYWORDS GI:24735005

SOURCE
 ORGANISM EST.
 Homo sapiens (human)

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 1 (bases 1 to 555)

TITLE
 Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
 Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
 Expressed sequence tag analysis of human RPE/choroid for the
 NEIBank Project: Over 6000 non-redundant transcripts, novel genes
 and splice variants
 Mol. Vis. 8 (4), 205-220 (2002)

JOURNAL
 MEDLINE
 PUBMED
 COMMENT 12107410
 Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0076
 Email: graeme@helix.nih.gov
 Plate: 92 row: h column: 10
 Seg primer: M13RPI reverse primer (ABI).
 Location/Qualifiers

FEATURES
 source
 1..555
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="cs92h10"
 /tissue_type="RPE/choroid"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Human Retinal pigment epithelium/choroid cDNA
 (Un-normalized, unamplified): cs"
 /note="Organ: Eye; Vector: pCMVSPORT6; Two different donor
 eyes (75-80 years old) yielded approximately 600 mg of
 dissected RPE/choroid tissue. This in turn yielded 340 ug
 of total RNA and 7 ug of mRNA. A directionally cloned cDNA
 library in the pCMVSPORT6 vector was constructed at Life
 Technologies (Rockville, MD; now part of Invitrogen Corp),
 essentially following the protocols of the Superscript
 Plasmid System (Invitrogen Corp).
 <http://www.invitrogen.com/> The library code
 designation was cs. For this library, cDNA inserts were
 cloned into the NotI/Mui sites of the vector. EST
 analysis was performed on the unamplified library at the
 NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:
 Pred. No.:

0.000319

Length:

555

Score: 135.00
 Percent Similarity: 82.76%
 Best Local Similarity: 82.76%
 Query Match: 79.88%
 DB: 14
 Gaps: 0
 Matches: 24
 Conservative: 0
 Mismatches: 5
 Indels: 0

US-10-010-050a-2_COPY_1_30 (1-30) x CA397574 (1-555)

QY 1 MetAArgAGlYAlaGlyAlaAlaArgAlaSerTrpCySerTrpAlaLeuAlaLeu 20
 Db 40 ATCGCGCGGGCGCGCGCGGT-CGCGGAGCGCGCTTCTGCTGCTGCGCGCTG 99
 QY 21 LeuTrpLeuAlaValAlaProGlyTrp 29
 Db 100 CTTTGCTCGCGGTGCTCGGCGCTGCTC 126

Search completed: April 25, 2004, 05:46:22
 Job time : 383.792 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 23:04:07 ; Search time 59.7917 Seconds

(without alignments)
2262.168 Million cell updates/sec

Title: US-10-010-050A-2_COPY_1_30
Perfect score: 169
Sequence: 1 MRGAGARGARASCMALALMLAVPGWS 30

Scoring table:
BIOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-Q=/cgn2_1/USPRO.spool/US10010050/funat_22042004_113205_27584/app_query.fasta_1.1372
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -MATRIX=bloname2
-TRANS=human40.cdd -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10010050@cgn_1_1.776@funat_22042004_113205_27584
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10C_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	169	100.0	1486	9	US-09-122-383-1	Sequence 1, Appl
2	169	100.0	1486	14	US-10-010-050A-1	Sequence 1, Appl
3	169	100.0	1751	13	US-10-653-595-26	Sequence 26, Appl
4	169	100.0	1751	13	US-09-397-945-26	Sequence 26, Appl
5	169	100.0	1751	13	US-09-122-383-13	Sequence 13, Appl
6	117	69.2	1038	9	US-09-122-383-13	Sequence 13, Appl
7	71	42.0	580	13	US-10-010-050A-13	Sequence 13, Appl
8	71	42.0	580	16	US-10-027-632-24716	Sequence 24716, Appl
9	71	42.0	580	13	US-10-027-632-24716	Sequence 24716, Appl
10	70.5	41.7	256	9	US-10-282-122A-15190	Sequence 15190, A
11	69.5	41.1	403	9	US-09-960-352-12900	Sequence 12900, A
12	69.5	41.1	403	9	US-09-960-352-12900	Sequence 12900, A
13	69.5	41.1	404	9	US-09-960-352-11275	Sequence 11275, A
14	68.5	40.5	4853	10	US-09-814-353-21894	Sequence 21894, A
15	68	40.2	4444	16	US-10-395-740-15	Sequence 15, Appl
16	67.5	39.9	276276	13	US-10-087-192-754	Sequence 754, Appl
17	66.5	39.3	3738	15	US-10-029-386-20677	Sequence 20677, A
18	66.5	39.3	3857	13	US-10-087-192-2057	Sequence 2057, A
19	66.5	39.3	24990	13	US-10-087-192-2056	Sequence 2056, A
20	66	39.1	7155	15	US-10-329-079-14	Sequence 14, Appl
21	66	39.1	37360	15	US-10-329-079-6	Sequence 6, Appl
22	65	38.5	468	13	US-10-027-632-181138	Sequence 181138, A
23	65	38.5	468	13	US-10-027-632-181139	Sequence 181139, A
24	65	38.5	468	16	US-10-027-632-181138	Sequence 181138, A
25	65	38.5	468	16	US-10-027-632-181139	Sequence 181139, A
26	64.5	38.2	12951	13	US-10-282-122A-31678	Sequence 31678, A
27	64	37.9	1490	13	US-10-280-576-25	Sequence 25, Appl
28	64	37.9	1530	13	US-10-282-122A-30126	Sequence 30126, A
29	64	37.9	2000	16	US-10-260-238-1708	Sequence 1708, A
30	63.5	37.6	594	9	US-09-950-933A-10	Sequence 10, Appl
31	63.5	37.6	1704	15	US-10-156-761-1987	Sequence 1987, A
32	63.5	37.6	9025608	15	US-10-156-761-1	Sequence 1, Appl
33	63	37.3	646	13	US-10-027-632-103173	Sequence 103173, A
34	63	37.3	646	16	US-10-027-632-103173	Sequence 103173, A
35	63	37.3	772	15	US-10-029-517-106	Sequence 106, Appl
36	63	37.3	1257	13	US-10-425-114-5227	Sequence 5227, A
37	63	37.3	96599	12	US-09-997-722-262	Sequence 262, Appl
38	63	37.3	332101	13	US-10-354-247-1	Sequence 1, Appl
39	63	37.3	322101	15	US-10-060-902-1	Sequence 1, Appl
40	62.5	37.0	1872	10	US-09-764-891-7060	Sequence 7060, A
41	62.5	37.0	1962	13	US-10-411-037-65	Sequence 65, Appl
42	62.5	37.0	1962	13	US-10-411-026-65	Sequence 299, Appl
43	62.5	37.0	2155	9	US-09-962-436-299	Sequence 2394, A
44	62.5	37.0	2155	9	US-09-880-107-2394	Sequence 1, Appl
45	62.5	37.0	6200	9	US-09-993-038-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-122-383-1
Sequence 1, Application US/09122383A
Patent No. US20020042093A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: CHROMOSOME 13
CURRENT APPLICATION NUMBER: US/09/122,383A
CURRENT FILING DATE: 1998-07-24
EARLIER APPLICATION NUMBER: 60/053,613
EARLIER FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1486
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (47)...(1084)

US-09-122-383-1

Alignment Scores:

Pred. No.:	4.18e-12	Length:	1486
Score:	169.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-122-383-1 (1-1486)

QY 1 MetAArgAGlYAlaGlyAlaAlaArgAGlYArgAlaSerTPCySTrPaAlaLeuAlaLeu 20

Db 47 ATGCGGGCGGGCGGGCGGGCGGGCTCGGGAGACGGCTTCTGTGCTGAGCGGCGTGGCGCTG 106

QY 21 LeuTPLeuAlaValValProGlyTTPSer 30

Db 107 CTTGGCTCGCGGTGTGTTCCGGGCTGTGTC 136

RESULT 2

US-10-010-050A-1

Sequence 1, Application US/10010050A
Publication No. US20020173624A1

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38C1
CURRENT APPLICATION NUMBER: US/10/010,050A
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 09/122,383
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 60/053,613
PRIOR FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1
LENGTH: 1486
TYPE: DNA
ORGANISM: Homo sapien

FEATURE:
NAME/KEY: CDS
LOCATION: (47)...(1084)

US-10-010-050A-1

Alignment Scores:

Pred. No.:	4.18e-12	Length:	1486
Score:	169.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050A-2_COPY_1_30 (1-30) x US-10-010-050A-1 (1-1486)

QY 1 MetAArgAGlYAlaGlyAlaAlaArgAGlYArgAlaSerTPCySTrPaAlaLeuAlaLeu 20

Db 47 ATGCGGGCGGGCGGGCGGGCGGGCTCGGGAGACGGCTTCTGTGCTGAGCGGCGTGGCGCTG 106

QY 21 LeuTPLeuAlaValValProGlyTTPSer 30

Db 107 CTTGGCTCGCGGTGTGTTCCGGGCTGTGTC 136

RESULT 3

US-10-653-595-26

Sequence 26, Application US/10653595
Publication No. US20040048304A1

GENERAL INFORMATION:

APPLICANT: Ruben et. al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027P1C1

CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03

PRIOR APPLICATION NUMBER: US 09/397945

PRIOR FILING DATE: 1999-09-17

PRIOR APPLICATION NUMBER: PCT/US99/05804

PRIOR FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: 60/078,566

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,576

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,573

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,574

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,579

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/080,314

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/080,312

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/078,578

PRIOR FILING DATE: 1998-03-19

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 470

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 26

LENGTH: 1751

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (1520)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1557)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1689)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1729)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1735)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1741)

OTHER INFORMATION: n equals a,t,g, or c

US-10-653-595-26

Alignment Scores:

Pred. No.:	4.93e-12	Length:	1751
Score:	169.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-10-010-050A-2_COPY_1_30 (1-30) x US-10-653-595-26 (1-1751)

QY 1 MetAArgAGlYAlaGlyAlaAlaArgAGlYArgAlaSerTPCySTrPaAlaLeuAlaLeu 20

Db 49 ATGCGGGCGGGCGGGCGGGCGGGCTCGGGAGACGGCTTCTGTGCTGAGCGGCGTGGCGCTG 108

QY 21 LeuTPLeuAlaValValProGlyTTPSer 30

Db 109 CTTGGCTCGCGGTGTGTTCCGGGCTGTGTC 138

RESULT 4		
US-09-397-945-26		
Sequence: 26, Application US/09397945		
Publication No. US20030065139a1		
GENERAL INFORMATION:		
APPLICANT: Human Genome Sciences, Inc. et al.		
TITLE OF INVENTION: 95 Human secreted proteins		
FILE REFERENCE: P20271		
CURRENT APPLICATION NUMBER: US/09/397,945		
CURRENT FILING DATE: 1999-09-17		
PRIOR APPLICATION NUMBER: PCT/US99/05804		
PRIOR FILING DATE: 1999-03-18		
PRIOR APPLICATION NUMBER: 60/078,566		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,576		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,573		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,574		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,579		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/080,314		
PRIOR FILING DATE: 1998-04-01		
PRIOR APPLICATION NUMBER: 60/080,312		
PRIOR FILING DATE: 1998-04-01		
PRIOR APPLICATION NUMBER: 60/078,578		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,581		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,577		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,563		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/080,313		
PRIOR FILING DATE: 1998-04-01		
NUMBER OF SEQ ID NOS: 470		
SOFTWARE: PatentIn Ver. 2.0		
SEQ ID NO: 26		
LENGTH: 1751		
TYPE: DNA		
ORGANISM: Homo sapiens		
FEATURE:		
NAME/KEY: SITE		
LOCATION: (1520)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1557)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1689)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1729)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1735)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1741)		
OTHER INFORMATION: n equals a,t,g, or c		
US-09-397-945-26		
Alignment Scores:		
Pred. No.:	4,93e-12	Length: 1751
Score:	169.00	Matches: 30
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	13	Gaps: 0
US-10-010-050A-2 COPY 1_30 (1-30) x US-09-397-945-26 (1-1751)		

```

QY      1 MetAArgAGGlyAlaGlyAlaAlaAlaArgGlyAArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
Db      49 ATGCGCGGCGGCGCGCGCGCGCGCTCGGCGGACGCGCTTCTGCTGCTGCTGCGCTGCGCTG 108
      21 LeuTrpLeuAlaValValProGlyTrpSer 30
Db      109 CTTTGGCTCGCGGTGCTTCCGGGCTGTGCTC 138

RESULT 5
US-09-122-383-13
; Sequence 13, Application US/09122383A
; Patent No. US20020042093A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; FILE REFERENCE: 97-38
; CURRENT APPLICATION NUMBER: US/09/122,383A
; CURRENT FILING DATE: 1998-07-24
; EARLIER APPLICATION NUMBER: 60/053,613
; EARLIER FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate nucleotide sequence encoding zsig46
; OTHER INFORMATION: polypeptide of SEQ ID NO:2
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)..(1038)
; OTHER INFORMATION: n is any nucleotide
US-09-122-383-13

Alignment Scores:
Pred. No.:          9.88e-06          Length:          1038
Score:              117.00            Matches:          20
Percent Similarity: 68.97%            Conservative:     0
Best Local Similarity: 68.97%          Mismatches:       9
Query Match:        69.23%            Indels:           0
Db:                  9                Gaps:              0

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-122-383-13 (1-1038)
QY      1 MetAArgAGGlyAlaGlyAlaAlaAlaArgGlyAArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
Db      1 ATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60
      21 LeuTrpLeuAlaValValProGlyTrp 29
Db      61 YTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 87

RESULT 6
US-10-010-050A-13
; Sequence 13, Application US/10010050A
; Publication No. US20020173624A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; FILE REFERENCE: 97-38C1
; CURRENT APPLICATION NUMBER: US/10/010,050A
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 09/122,383
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/053,613
; PRIOR FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0

```



```
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 11275
LENGTH: 404
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 48-LIB34-036-Q1-E1-D8
US-09-960-352-11275

Alignment Scores:
Pred. No.: 3.54 Length: 404
Score: 69.50 Matches: 15
Percent Similarity: 48.48% Conservative: 1
Best Local Similarity: 45.45% Mismatches: 10
Query Match: 41.12% Indels: 7
DB: 9 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-960-352-11275 (1-404)
QY 4 G1yAlaG1yAlaAlaArgG1yArgAlaSerTyrCys-----Trr 16
DB 304 GGACGAGGTGCATACGCGTCCGTGTCATGTTGCCGTACTTGAGCCGTTTCATGG 245
QY 17 AlaleuAlaleuLeuTrrPleuAlaValProGlyTrr 29
DB 244 GCTCTGGGCGCTGCATCGGCGGCGCCGCATCAGCATGG 206

RESULT 13
US-09-960-352-14095/C
Sequence 14095, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Mengping
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 14095
LENGTH: 414
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 60-LIB34-005-Q1-E1-G8
US-09-960-352-14095

Alignment Scores:
Pred. No.: 3.63 Length: 414
Score: 69.50 Matches: 15
Percent Similarity: 48.48% Conservative: 1
Best Local Similarity: 45.45% Mismatches: 10
Query Match: 41.12% Indels: 7
DB: 9 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-960-352-14095 (1-414)
QY 4 G1yAlaG1yAlaAlaArgG1yArgAlaSerTyrCys-----Trr 16
DB 402 GGACGAGGTGCATACGCGTCCGTGTCATGTTGCCGTACTTGAGCCGTTTCATGG 343
QY 17 AlaleuAlaleuLeuTrrPleuAlaValProGlyTrr 29
DB 342 GCTCTGGGCGCTGCATCGGCGGCGCCGCATCAGCATGG 304

RESULT 14
US-09-814-353-21894/C
Sequence 21894, Application US/09814353
Publication No. US20030165831A1
```

```
GENERAL INFORMATION:
APPLICANT: Lee, John
APPLICANT: Thompson, Pamela
APPLICANT: Lillie, James
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF OVARIAN CANCER
FILE REFERENCE: MRI-0063
CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21894
LENGTH: 4853
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1, 2, 3, 4, 4827, 4828, 4829, 4830, 4831, 4832, 4833, 4834,
LOCATION: 4835, 4836, 4837, 4838, 4839, 4840, 4841, 4842, 4843, 4844,
LOCATION: 4845, 4846, 4847, 4848, 4849, 4850, 4851, 4852, 4853
OTHER INFORMATION: n = A, T, C or G
US-09-814-353-21894

Alignment Scores:
Pred. No.: 56.5 Length: 4853
Score: 68.50 Matches: 14
Percent Similarity: 62.07% Conservative: 4
Best Local Similarity: 48.28% Mismatches: 10
Query Match: 40.53% Indels: 1
DB: 10 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-814-353-21894 (1-4853)
QY 1 MetArgArgG1yAlaG1yAlaAlaArgG1yArgAlaSerTyrCysTrrPalaLeuAlaLeu 20
DB 4590 ATTGGGCGGCTTGTGTGGTGGCGCC--GGGGGGGTGTGTTTGTCTGGTGGTATTGG 4534
QY 21 LeuTrrPleuAlaValProGlyTrr 29
DB 4533 TGCTGGAGTGGCGGTCTTGGCCGCTTG 4507

RESULT 15
US-10-395-740-15
Sequence 15, Application US/10395740
Publication No. US20030215852A1
GENERAL INFORMATION:
APPLICANT: WYETH
APPLICANT: Mark, Robert
APPLICANT: Wood, Andrew
APPLICANT: Gulukota, Kamalakr
TITLE OF INVENTION: NOVEL PANCREATIN-PABLO PROTEIN INTERACTIONS AND METHODS OF USE
TITLE OF INVENTION: THEREOF
FILE REFERENCE: AM100375PCT
CURRENT APPLICATION NUMBER: US/10/395,740
CURRENT FILING DATE: 2003-03-24
PRIOR APPLICATION NUMBER: US 60/369,244
PRIOR FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: US 60/386,645
PRIOR FILING DATE: 2002-06-06
NUMBER OF SEQ ID NOS: 15
```

SOFTWARE: PatentIn version 3.2

SEQ ID NO 15

LENGTH: 44442

TYPE: DNA

ORGANISM: Genomic DNA

IS-10-395-740-15

Alignment Scores:

Prod. No.:	595	Length:	44442
Score:	68.00	Matches:	13
Percent Similarity:	53.12%	Conservative:	4
Best Local Similarity:	40.62%	Mismatches:	13
Query Match:	40.24%	Indels:	2
DB:	16	Gaps:	1

IS-10-010-050A-2_COPY_1_30 (1-30) x US-10-395-740-15 (1-44442)

21 MetArgArgGlyValIaGlyIaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20

27846 ATGCAAGAGAGGAAAGAGGAGCCCTTGAGCCACGCGAGGTTGGGCTTGGGGCAGAGCCGCC 27905

21 LeuTrpLeuAlaValVal-----ProGlyTrpSer 30

27906 CTCTGAGAGCAAGACTGCAAGGTCCAGGTTGAGCG 27941

Search completed: April 25, 2004, 06:14:58
Job time : 68.7917 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - nucleic search, using frame_plus_p2n model

on: April 24, 2004, 23:02:52 ; Search time 9.95833 Seconds

(without alignment)
1671.819 Million cell updates/sec

file: US-10-010-050A-2_COPY_1_30

target score: 169

sequence: 1 MRGAGARGRASWCALALMLAVPGMS 30

scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 682709 seqs, 277475446 residues

total number of hits satisfying chosen parameters: 1365418

minimum DB seq length: 0
maximum DB seq length: 2000000000

hit-processing: Minimum Match 0%

Maximum Match 100%

listing first 45 summaries

command line parameters:

CODE=frame_p2n.model -DEV=x1h
-p=/cg2_1/USPTO/spool/US10010050/runat_22042004_113205_27569/app_query.fasta_1.1372
-B=Issued Patents NA -QFMT=fastp -SUFFIX=rm -MINMATCH=0.1 -LOOFCU=0
-OOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-IST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-CODE=LOCAL -OUTFMT=pco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-JER=US10010050 @CGN_1_1_140 @runat_22042004_113205_27569 -NCPU=6 -ICPU=3
-B MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOC
-JVT TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

itabase: Issued Patents NA:*

1: /cg2_6/pdata/2/ina/5A.COMB.seq:*
2: /cg2_6/pdata/2/ina/5B.COMB.seq:*
3: /cg2_6/pdata/2/ina/6A.COMB.seq:*
4: /cg2_6/pdata/2/ina/6B.COMB.seq:*
5: /cg2_6/pdata/2/ina/PCTUS.COMB.seq:*
6: /cg2_6/pdata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	64	37.9	438	4	US-09-252-991A-9484
2	64	37.9	594	4	US-09-252-991A-7647
3	64	37.9	1533	4	US-09-252-991A-9378
4	64	37.9	2043	4	US-09-252-991A-9459
5	64	37.9	2274	4	US-09-252-991A-9347
6	63.5	37.6	2316	4	US-09-252-991A-13835
7	63.5	37.6	2811	4	US-09-252-991A-13619
8	63	37.3	700	1	US-08-037-579A-4
9	63	37.3	700	3	US-08-601-184-4
10	62.5	37.0	2067	2	US-08-713-928B-8
11	62.5	37.0	2155	3	US-09-191-171-4
12	62.5	37.0	2155	3	US-09-385-707-4

C 13	62.5	37.0	2155	4	US-09-639-696C-4	Sequence 4, Appli
C 14	62.5	37.0	4480	3	US-09-191-171-7	Sequence 7, Appli
C 15	62.5	37.0	4480	3	US-09-385-707-7	Sequence 7, Appli
C 16	62.5	37.0	6200	4	US-09-439-923-1	Sequence 1, Appli
C 17	62.5	37.0	6200	4	US-09-711-202A-1	Sequence 1, Appli
C 18	62.5	37.0	6200	4	US-09-711-205A-1	Sequence 1, Appli
C 19	62.5	37.0	6238	4	US-09-639-696C-6	Sequence 6, Appli
C 20	62.5	37.0	169998	4	US-09-676-610B-24	Sequence 24, Appli
C 21	62.5	37.0	17496	4	US-09-877-177A-10	Sequence 10, Appli
C 22	62	36.7	365	4	US-09-621-976-11459	Sequence 11459, A
C 23	62	36.7	825	4	US-09-489-039A-5870	Sequence 5870, Ap
C 24	62	36.7	2502	4	US-09-489-039A-1765	Sequence 1765, Ap
C 25	62	36.7	3780	4	US-09-489-039A-1669	Sequence 1669, Ap
C 26	61.5	36.4	702	4	US-09-252-991A-1115	Sequence 1115, Ap
C 27	61.5	36.4	2178	4	US-09-252-991A-1039	Sequence 1039, Ap
C 28	61.5	36.4	2433	4	US-09-252-991A-923	Sequence 923, App
C 29	61	36.1	972	4	US-09-489-039A-2993	Sequence 2993, Ap
C 30	61	36.1	2502	4	US-09-069-023-2	Sequence 2, Appli
C 31	60.5	35.8	657	4	US-09-489-039A-2986	Sequence 2986, Ap
C 32	60.5	35.8	1452	4	US-09-489-039A-2735	Sequence 2735, Ap
C 33	60.5	35.8	43280	2	US-08-804-227C-1	Sequence 5631, Ap
C 34	60.5	35.5	894	4	US-09-489-039A-5631	Sequence 5264, Ap
C 35	60	35.5	1113	4	US-09-489-039A-5264	Sequence 4, Appli
C 36	60	35.5	1208	2	US-08-403-852D-4	Sequence 4, Appli
C 37	60	35.5	1208	3	US-08-510-646B-4	Sequence 4, Appli
C 38	60	35.5	1208	3	US-09-231-818-4	Sequence 4, Appli
C 39	60	35.5	1208	4	US-09-635-359B-4	Sequence 13, Appli
C 40	60	35.5	1611	4	US-08-808-701A-13	Sequence 1, Appli
C 41	60	35.5	2126	2	US-08-789-354-1	Sequence 1, Appli
C 42	60	35.5	2126	3	US-09-110-937-1	Sequence 1, Appli
C 43	60	35.5	2126	3	US-09-058-725B-1	Sequence 1, Appli
C 44	60	35.5	2126	3	US-09-232-857-1	Sequence 1, Appli
C 45	60	35.5	5392	2	US-08-403-852D-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-9484
Sequence 9484, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubinfeld et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 9484
LENGTH: 438
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9484

Alignment Scores:

Pred. No.: 33.2
Score: 64.00
Percent Similarity: 45.24%
Best Local Similarity: 40.48%
Query Match: 37.87%
DB: 4
Gaps: 3

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-252-991A-9484 (1-438)

Oy 2 ArgArgGlyAlaGlyAlaAlaArgGlyAlaSerTP-----14
DB 5 CGCGACCCGCTGGCGGCGCTCGCAGCGGCGGCGATGCGATGATCGACGACGACGAGG 64

```

Qy      15 -----Cys-----TPPalaIeuAlaIeuTrpIeu-----AlaValPro 27
          |||              |||::||| ||| ||| ||| ||| ||| ||| |||
Db       65 TATGCTGGAGCAGATGTGCTCCGGCGGCACCTAATGCTTCAGCCGGCGCAGCAATTCT 124

Qy      28 GlyTrp 29
          |||
Db       125 GCTTGG 130

RESULT 2
US-09-252-991A-7647/c
Sequence 7647, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7647
LENGTH: 594
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7647

Alignment Scores:
Pred. No.:           46.4             Length:         594
Score:               64.00            Matches:         15
Percent Similarity:  47.92%           Conservative:    8
Best local similarity: 31.25%          Mismatches:     7
Query Match:         37.87%           Indels:         18
DB:                  4                Gaps:           3

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-252-991A-7647 (1-594)

Qy      1 MetArgArgGlyAlaGlyAlaAlaAlaArgGlyArg----- 1..
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       180 GTGGAGCGGGCGGCTGGCGGGCGCAAGGGTGACGGGATGACTGCTGTGAGGTTCC 121

Qy      12 AlaSerTrpCysTrpAlaIeuAlaIeuIeu-----TrpIeu----- 23
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       120 TGCACTGGCGGCTCGAGGGTGTTGATGTGGCATCACTGGGTGGGAAGCGTCGAAC 6..

Qy      24 ---AlaValValProGlyTrnSer 30
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       60 TCGGCGGTATGGCGGCCCTGGGCC 37

RESULT 3
US-09-252-991A-9378/c
Sequence 9378, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 9378
LENGTH: 1533
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9378
```

```

Alignment Scores:
Pred. No.: 131 Length: 1533
Score: 64.00 Matches: 17
Percent Similarity: 45.24% Conservative: 2
Best Local Similarity: 40.48% Mismatches: 9
Query Match: 37.87% Indels: 14
DB: 4 Gaps: 3

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-252-991A-9378 (1-1533)

QY 2 ArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrp-----14
Db 418 CGGGGACCCGCTGGGGCCGCTCGCAGCGGCGGAGATGAGGTGATGACAGCAGCAGG 359
QY 15 ---Cys-----TrpAlaLeuAlaLeuLeuTrpLeu-----AlaValValPro 27
Db 358 TATGTCTGGCGCAGAGATGATCTCGGGCCGACATATGAGCTTCCAGCCGGCGAGCAATTCT 299
QY 28 GlyTrp 29
Db 298 GCTTGG 293

RESULT 4
US-09-252-991A-9459
; Sequence 9459, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9459
; LENGTH: 2043
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9459

Alignment Scores:
Pred. No.: 180 Length: 2043
Score: 64.00 Matches: 17
Percent Similarity: 45.24% Conservative: 2
Best Local Similarity: 40.48% Mismatches: 9
Query Match: 37.87% Indels: 14
DB: 4 Gaps: 3

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-252-991A-9459 (1-2043)

QY 2 ArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrp-----14
Db 777 CGGGGACCCGCTGGGGCCGCTCGCAGCGGCGGAGATGAGGTGATGACAGCAGCAGG 836
QY 15 ---Cys-----TrpAlaLeuAlaLeuLeuTrpLeu-----AlaValValPro 27
Db 837 TATGTCTGGCGCAGAGATGATCTCGGGCCGACATATGAGCTTCCAGCCGGCGAGCAATTCT 896
QY 28 GlyTrp 29
Db 897 GCTTGG 902

RESULT 5
US-09-252-991A-9347/c
; Sequence 9347, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```


NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7956-0011-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2067 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
IS-08-713-928B-8

Alignment Scores:
Pred. No.: 268 Length: 2067
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 36.98% Indels: 11
DB: 2 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x US-08-713-928B-8 (1-2067)

Y 2 ArgArgG1yAla-----G1yAlaAlaArgG1y 10
b 1081 CGAAGGGGTGCGGGGTGAGTCTCAGAGGCAATTGCTGTTGCTCAGAGGCGCGTAGGGGA 1022

Y 11 ArgAlaSerTTPCySTTPAlaLeuAla 19
b 1021 AGGCGGAGGTGGTGTGTTGGCCAGTAGCA 995

RESULT 11
S-09-191-171-4/c
Sequence 4, Application US/09191171
Patent No. 6149909
Patent No. 6143294
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
TITLE OF INVENTION: SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,171
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00611

FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 89..2047
US-09-191-171-4

Alignment Scores:
Pred. No.: 281 Length: 2155
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 36.98% Indels: 11
DB: 3 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-191-171-4 (1-2155)

QY 2 ArgArgG1yAla-----G1yAlaAlaArgG1y 10
Db 1169 CGAAGGGGTGCGGGGTGAGTCTCAGAGGCAATTGCTGTTGCTCAGAGGCGCGTAGGGGA 1110

QY 11 ArgAlaSerTTPCySTTPAlaLeuAla 19
Db 1109 AGGCGGAGGTGGTGTGTTGGCCAGTAGCA 1083

RESULT 12
US-09-385-707-4/c
Sequence 4, Application US/09385707
Patent No. 6238662
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
TITLE OF INVENTION: SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,707
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,104
FILING DATE: 13-NOV-1998
ATTORNEY/AGENT INFORMATION:
NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 89..2047
US-09-385-707-4

Alignment Scores:
Pred. No.: 281 Length: 2155
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 36.98% Indels: 11
DB: 3 Gaps: 1

US-10-010-050a-2_copy_1_30 (1-30) x US-09-385-707-4 (1-2155)
QY 2 ArgAArgGlyA-----GlyAlaAlaArgGly 10
DB 1169 CGAAGGGGGTGGGGTGTAGCTCAGGAAGCATTTGCTTGCCTCAGAGGCGCGTAGGGGA 1110

QY 11 ArgAlaSerTPCySTPAlaLeuAla 19
DB 1109 AGGCGGAGGTGTGTGTGCCAGTAGCA 1083

RESULT 13
US-09-639-696C-4/c
Sequence 4, Application US/09639696C
Patent No. 6524835
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
ANSON, Donald S.
ORSBORN, Annette M.
NELSON, Paul V.
CLEMENS, Peter R.
MORRIS, Charles P.
HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON PEABODY LLP
STREET: 990 Stewart Avenue
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/639,696C
FILING DATE: 16-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/191,171
FILING DATE: 13-NOV-1998
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pokalsky, Ann R.
REGISTRATION NUMBER: 34,697
REFERENCE/DOCKET NUMBER: 2249/304
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-832-7572
TELEFAX: 516-832-7555
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 89..2047
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-639-696C-4

Alignment Scores:
Pred. No.: 281 Length: 2155
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 36.98% Indels: 11
DB: 4 Gaps: 1

US-10-010-050a-2_copy_1_30 (1-30) x US-09-639-696C-4 (1-2155)
QY 2 ArgAArgGlyA-----GlyAlaAlaArgGly 10
DB 1169 CGAAGGGGGTGGGGTGTAGCTCAGGAAGCATTTGCTTGCCTCAGAGGCGCGTAGGGGA 1110

QY 11 ArgAlaSerTPCySTPAlaLeuAla 19
DB 1109 AGGCGGAGGTGTGTGTGCCAGTAGCA 1083

RESULT 14
US-09-191-171-7/c
Sequence 7, Application US/09191171
Patent No. 6149909
Patent No. 6149909 6143294
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
ANSON, Donald S.
ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,171
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8978Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-191-171-7

Alignment Scores:
Pred. No.: 627 Length: 4480
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 36.98% Indels: 11
Gaps: 1

US-10-010-050a-2_COPY_1_30 (1-30) x US-09-191-171-7 (1-4480)

Y 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
|||
b 2192 CGAAGGGGTGGCGGTGAGCTCAGGAGGCAATTGCTGCTCAGAGCGCGGTAGGGA 2133
|||
Y 11 ArgAlaSerTTPCyETTPAlaLeuAla 19
|||
b 2132 AGGCGAGGTGGTGTGGCCAGTAGCA 2106
|||

RESULT 15
US-09-385-707-7/c
Sequence 7, Application US/09385707
Patent No. 6238662
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
TITLE OF INVENTION: SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York

COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,707
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,104
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8978Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-385-707-7

Alignment Scores:
Pred. No.: 627 Length: 4480
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 36.98% Indels: 11
Gaps: 1

US-10-010-050a-2_COPY_1_30 (1-30) x US-09-385-707-7 (1-4480)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
|||
Db 2192 CGAAGGGGTGGCGGTGAGCTCAGGAGGCAATTGCTGCTCAGAGCGCGGTAGGGA 2133
|||
QY 11 ArgAlaSerTTPCyETTPAlaLeuAla 19
|||
Db 2132 AGGCGAGGTGGTGTGGCCAGTAGCA 2106
|||

Search completed: April 25, 2004, 05:51:41
Job time: 12.9583 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 22:59:01 ; (Search time 55.375 Seconds
(without alignments)
2301.507 Million cell updates/sec

Title: US-10-010-050a-2_COPY_1_30

Perfect score: 169
Sequence: 1 MRRGAGARGRGRASWCWALALMLAVVPGMS 30

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODE=frame+ p2n.model -DEV=x1h
Q=/cgr2_1/USFTO.spool/US10010050/runat.22042004.113203.27539/app.query.fasta.1.1372
DB=N_Geneseq.250a04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdi
LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=pct -NORM=0 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US10010050 @CGN 1.1.819 @runat.22042004.113203.27539 -NCPD=6 -ICPD=3
NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

N_Geneseq.250a04:*
1: geneseqn1808:*
2: geneseqn1908:*
3: geneseqn2000:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	169	100.0	697	4	AAK93424 Human cDN
2	169	100.0	697	4	AAK92158 Human cDN
3	169	100.0	1486	2	AAK02855 Human zfi
4	169	100.0	1751	2	AAZ24826 Human sec
5	169	100.0	2120	4	AAK94829 Human ful
6	131	77.5	630	6	ABQ40653 Oligonuc1
7	131	77.5	630	6	ABQ40652 Oligonuc1
8	131	77.5	631	6	ABQ13702 Oligonuc1

C	9	131	77.5	631	6	ABQ13703	Abq13703 Oligonuc1
C	10	117	69.2	1038	2	AAK02866	Aax02866 Human deg
C	11	71	42.0	1071	7	ACA27320	Aca27320 Prokaryot
C	12	70.5	41.7	256	7	ABX47735	Abx47735 Bovine ES
C	13	70	41.4	882	9	ADCO8897	Adc08897 Rice DNA
C	14	69.5	41.1	403	7	ABX40924	Abx40924 Bovine ES
C	15	69.5	41.1	404	7	ABX46110	Abx46110 Bovine ES
C	16	69.5	41.1	414	7	ABX48930	Abx48930 Bovine ES
C	17	69	40.8	2028	2	AAK54769	Aax54769 Endotheli
C	18	69	40.8	2028	3	AAA43426	Aax43426 Human ade
C	19	69	40.8	2028	3	AAF20338	Aaf20338 Human end
C	20	69	40.8	2028	7	ABZ96032	Abz96032 Human end
C	21	69	40.8	6225	2	AAK55273	Aax55273 Human enz
C	22	69	40.8	6225	3	AAA34721	Aa34721 Human ade
C	23	69	40.8	6225	3	AAF20843	Aaf20843 Human mul
C	24	69	40.8	6225	7	ABZ96537	Abz96537 Human nuc
C	25	69	40.8	35384	3	AAF21436	Aaf21436 Human enz
C	26	69	40.8	35459	7	ABZ97130	Abz97130 Human nuc
C	27	69	40.8	114955	2	AAK53431	Aax53431 Human nuc
C	28	68	40.2	44442	9	ADD08990	Aad08990 Human ade
C	29	67.5	39.9	5376	6	ABK84080	Abk84080 Human pan
C	30	67	39.6	98690	6	ABK12169	Abk12169 Human DNA
C	31	66.5	39.3	4848	5	AAK67953	Aas67953 DNA encod
C	32	66.5	39.3	5592	5	AAK67958	Aas67958 DNA encod
C	33	66	39.1	3318	4	ABL04056	AbL04056 Drosophi1
C	34	65	38.5	2706	5	AAK80870	Aas80870 DNA encod
C	35	64.5	38.2	12951	7	ACA43808	Aca43808 Prokaryot
C	36	64.5	38.2	58857	3	AAA58471	Aa58471 Nucleotid
C	37	64	37.9	1490	7	ACC57671	Acc57671 Mouse pro
C	38	64	37.9	1530	7	ACA42256	Aca42256 Prokaryot
C	39	64	37.9	2000	7	ADA71534	Ada71534 Rice gene
C	40	64	37.9	2853	4	ABL29393	AbL29393 Drosophi1
C	41	64	37.9	6906	4	ABL29392	AbL29392 Drosophi1
C	42	63.5	37.6	594	6	AAI37818	AaI37818 Wheat KCP
C	43	63	37.3	622	4	AAK74135	Aak74135 Human imm
C	44	63	37.3	698	9	ADQ30433	Adq30433 Human nov
C	45	63	37.3	700	2	AAQ76361	Aaq76361 Human imm

ALIGNMENTS

RESULT 1	AAK93424	standard; cDNA, 697 BP.
XX	AAK93424	
AC	AAK93424	
XX		
DT	06-NOV-2001	(first entry)
XX		
DE	Human cDNA clone representative sequence, SEQ ID NO: 1884.	
XX	Human, full length cDNA; cDNA synthesis; oligo-capping; ss.	
KW	Human, full length cDNA; cDNA synthesis; oligo-capping; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	EP1130094-A2.	
XX		
PD	05-SEP-2001.	
XX		
PF	07-JUL-2000; 2000EP-00114089.	
XX		
PR	08-JUL-1999; 99JP-00194486.	
PR	11-JAN-2000; 2000JP-00118774.	
PR	02-MAY-2000; 2000JP-00183765.	
XX		
XX	(HELI-) HELIX RES INST.	
PA		
XX		
PI	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,	
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX		
DR	WPI, 2001-52425/58.	
XX		
PT	830 Primers useful for synthesizing full length cDNA clones and their use	

XX EP1130094-A2.
XX
XX
PD 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-00114089.
PF
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Oka T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
XX
XX P-PSDB; AAM93870.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
XX
PS Claim 8; SEQ ID NO 3977; 1380bp + Sequence Listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO
XX
SQ Sequence 2120 BP; 590 A; 409 C; 460 G; 661 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
Pred. No.: 1.23e-09 Length: 2120
Score: 169.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
XX
US-10-010-050A-2_COPY_1_30 (1-30) x AAK94829 (1-2120)
XX
XX
XX 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
DB 52 ATGCGGGCGGGCGGGCGGGCGGGCTCGGGGACGGCCTTCTGTGCTGGGCGCTG 111
XX
XX 21 LeuTrpLeuAlaValValProGlyTrpSer 30
DB 112 CTTTGCGCTCGCGGCTTCGCGGCTCGCGGCTGCTGCTGCGGCGCTG 111
XX
XX
XX RESULT 6
ABQ40653/c
ID ABQ40653 standard; DNA; 630 BP.
XX
XX AC ABQ40653;
XX
XX 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 27244.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX
OS

XX
XX WO200218632-A2.
XX
XX
XX 07-MAR-2002.
XX
XX
XX 01-SEP-2001; 2001WO-EP010074.
XX
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX 05-SEP-2000; 2000DE-01044543.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX
XX Claim 12; 56bp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX
XX SQ Sequence 630 BP; 219 A; 255 C; 70 G; 86 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
Pred. No.: 6.04e-06 Length: 630
Score: 131.00 Matches: 22
Percent Similarity: 75.86% Conservative: 0
Best Local Similarity: 75.86% Mismatches: 7
Query Match: 77.51% Indels: 0
DB: 6 Gaps: 0
XX
US-10-010-050A-2_COPY_1_30 (1-30) x ABQ40653 (1-630)
XX
XX
XX 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
DB 332 ATGCGGGCGGGCGGGCGGGCGGGCTTCGGGGACGGCTTTTGTGCTGCGGCTG 273
XX
XX 21 LeuTrpLeuAlaValValProGlyTrp 29
DB 272 TTTTGCTCGCGGCTTCGCGGCTTCGCGGCTTCG 246
XX
XX
XX RESULT 7
ABQ40652
ID ABQ40652 standard; DNA; 630 BP.
XX
XX AC ABQ40652;
XX
XX 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 27243.
XX
XX
XX

Db 269 TTTTGGTTGGCGGAGTTCGGGTTGG 295

RESULT 9

ABQ13703/c

ID ABQ13703 standard; DNA, 631 BP.

XX ABQ13703;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 294.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP010074.

XX 01-SEP-2000; 2000DE-01043826.

XX 05-SEP-2000; 2000DE-01044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Gueutig D;

XX MPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNP's); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ54121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention

XX Sequence 631 BP; 220 A; 256 C; 70 G; 85 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.: 6.05e-06 Length: 631
XX Score: 131.00 Matches: 22
XX Percent Similarity: 75.86% Conservative: 0
XX Best Local Similarity: 75.86% Mismatches: 7
XX Query Match: 77.51% Indels: 0
XX Gaps: 0

JS-10-010-050a-2_COPY_1_30 (1-30) x ABQ13703 (1-631)

QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTyrCysTrpAlaLeuAlaLeu 20

Db 423 ATGGCGGCGGGCGCGCGCGGTTCCGGGAGCGCTTTTGGTGTGGCTTTGGCGTTG 364

QY 21 LeuTrpLeuAlaValAlaProGlyTyr 29

Db 363 TTTTGGTTGGCGGAGTTCGGGTTGG 337

RESULT 10

AAQ02865

ID AAQ02866 standard; DNA, 1038 BP.

XX AAQ02866;

XX 14-MAY-1999 (first entry)

XX Human degenerate zsig46 DNA.

XX Secreted protein; zsig46; human; chromosome 13; thyroid; disease;
XX hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;
XX Hirschsprung's disease; neuronal ceroid-lipofusiosis; Wilson disease;
XX Reiger syndrome; immunoassay; detection; anti-idiotypic antibody;
XX therapy; diagnostic; ss.

XX Homo sapiens.

XX WO9905275-A1.

XX 04-FEB-1999.

XX 24-JUL-1998; 98WO-US015431.

XX 24-JUL-1997; 97US-0053613P.

XX (ZYMO) ZYMOGENETICS INC.

XX Sheppard PO, Gilbertson DG;

XX MPI; 1999-142930/12.

PT New secreted polypeptide, zsig46, and its fragments, related fusion
PT proteins - used for diagnosis and treatment of thyroid disorders or
PT diseases involving genes on chromosome 13.

PS Claim 31; Page 94-95; 101pp; English.

XX This invention describes the isolation of a novel human secreted protein,
XX zsig46 encoded by a gene on chromosome 13 which is mainly expressed in
XX the thyroid. This product can be used to study secretion of proteins from
XX cells and also to treat or prevent deficient expression of zsig46, which
XX may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
XX disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
XX involve genes in the same region of chromosome 13 (e.g. Hirschsprung's
XX disease, neuronal ceroid-lipofusiosis, Wilson disease and Reiger
XX syndrome). Antibodies and other binding proteins, are used as immunoassay
XX reagents to detect zsig46 or cells expressing it, e.g. for assessing
XX thyroid function to produce anti-idiotypic antibodies, for affinity
XX purification of zsig46, to screen expression libraries, to neutralise
XX zsig46 activity, and to deliver toxins, radioisotopes etc. for
XX therapeutic or diagnostic purposes. Antigens of the product can be used
XX to promote growth, differentiation and proliferation of specific cell
XX types, e.g. for treating (extra)thyroid diseases or as additive to cell
XX cultures

XX Sequence 1038 BP; 189 A; 106 C; 179 G; 166 T; 0 U; 398 Other;

XX Alignment Scores:

XX Pred. No.: 0.0004 Length: 1038
XX Score: 117.00 Matches: 20
XX Percent Similarity: 68.97% Conservative: 0
XX Best Local Similarity: 68.97% Mismatches: 9
XX Query Match: 69.23% Indels: 0
XX Gaps: 0

CC cattle, and the LMPD nucleic acid can specifically hybridise to a second
CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC appearing as ABX34936-ABX49997, or complements of them. Also included are
CC ; (1) a transformed cell having a nucleic acid comprising an LMPD nucleic
CC acid linked to a promoter and a 3' non-translated sequence that
CC functions in the cell to cause termination of transcription and addition
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC (2) determining a level or pattern of a molecule in a bovine cell or
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC of the 15112 nucleic acid sequences or its complement or fragment) with a
CC complementary nucleic acid molecule obtained from the bovine cell or
CC tissue, where hybridisation between the marker nucleic acid and the
CC complementary nucleic acid permits the detection of the molecule; and (b)
CC detecting the level or pattern of the complementary nucleic acid, where
CC the detection of the complementary nucleic acid is predictive of the
CC level or pattern of the molecule. The LMPD nucleic acid is used for
CC determining a level or pattern of a molecule in a bovine cell or tissue.
CC It is useful for genome mapping, gene identification and analysis, cattle
CC breeding, preparation of constructs for use in cattle gene expression, or
CC for genetically improving cattle. The present sequence is one of the
CC 15112 bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The
CC present sequence was not shown in the specification but was obtained in
CC electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docID=20020137139

XX Sequence 256 BP; 62 A; 69 C; 75 G; 50 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 14.5 Length: 256
Score: 70.50 Matches: 15
Percent Similarity: 45.45% Conservative: 0
Best Local Similarity: 45.45% Mismatches: 11
Query Match: 41.72% Indels: 7
DB: Gaps: 1

JS-10-010-050a-2_COPY_1_30 (1-30) x ABX47735 (1-256)

2Y 4 GYALAGLYALALaAGGlyAGGAlaSerTTPys-----TTP 16
112 GGAGCAGAGTGCACACGCTGCTCATGTGCTCATGTCCTTCTTCATGG 53

2Y 17 ALaleuAlaleuLeuTTPleuAlaValaProGlyTTP 29
52 GCTCTGGGCGCTGCATGGGCGAGGCCGCCGATCAGCAGG 14

RESULT 13

ADC08897
ID ADC08897 standard; DNA; 882 BP.

AC ADC08897;

XX 18-DEC-2003 (first entry)

DE Rice DNA sequence seg ID323 related to grain filling.

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corr;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarcane;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW gene; ds; plant.

OS Oryza sativa.

XX WO200300905-A2.

XX 03-JAN-2003.

XX 21-JUN-2002; 2002WO-IB002450.

XX 22-JUN-2001; 2001US-0300112P.

XX 26-SEP-2001; 2001US-0325277P.

XX 20-DEC-2001; 2001US-0342327P.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;

PT Glaesbrook U, Katagiri F, Kleps J, Provatt N, Riche D;

XX WPI; 2003-229341/22.

DR P-PSDB; ADC08057.

PT New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.

PS Claim 30; SEQ ID NO 323; 130pp; English.

XX This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarcane, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence encoding a rice protein of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpct_sequences.

XX Sequence 882 BP; 175 A; 307 C; 260 G; 140 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 66.1 Length: 882
Score: 70.00 Matches: 12
Percent Similarity: 55.56% Conservative: 3
Best Local Similarity: 44.44% Mismatches: 12
Query Match: 41.42% Indels: 0
DB: Gaps: 0

US-10-010-050a-2_COPY_1_30 (1-30) x ADC08897 (1-882)

QY 2 ArgAArgLYALaGlyALaAlaArgGlyAGGAlaSerTTPysTTPAlaLeuAlaLeu 21
35 CGCGGGGTTTCAGTTTCACACCGAGGAGCAGAGCTGTGTGCTGACTACTCTTCAC 94

QY 22 TTPleuAlaValaProGly 28

DB 95 AGCTTGCAGCGCGGCGCGG 115

RESULT 14

ABX40924/c
ID ABX40924 standard; cDNA; 403 BP.

AC ABX40924;

XX 20-FEB-2003 (first entry)

DE Bovine EST associated with lactation/muscle/fat deposition #6089.

XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.

OS Bos Taurus.

XX US2002137139-A1.

XX 26-SEP-2002.

XX 24-SEP-2001; 2001US-00960352.

XX 12-JAN-1999; 99US-0115707P.

```

R      11-JAN-2000; 2000US-00480902.
A      (BYAT/) BYATT J C.
A      (MATH/) MATHIALAGAN N.
A      (TAON/) TAO N.
A      (WARR/) WARREN W C.
T      Byatt JC, Mathialagan N, Tao N, Warren WC;
T      MPI; 2003-110599/10.
T      New nucleic acid associated with lactation, and muscle and fat
T      deposition, useful for genome mapping, gene identification and analysis,
T      cattle breeding, or for genetically improving cattle.
X      Claim 2; SEQ ID NO 6089; 245bp; English.
C      The invention relates to a purified nucleic acid molecule associated with
C      lactation or muscle and fat deposition (designated LMPD), derived from
C      cattle, and the LMPD nucleic acid can specifically hybridise to a second
C      nucleic acid molecule comprising any of 15112 nucleotide sequences,
C      appearing as ABX34836-ABX49947, or complements of them. Also included are
C      (1) a transformed cell having a nucleic acid comprising an LMPD nucleic
C      acid linked to a promoter and a 3' non-translated sequence that
C      functions in the cell to cause termination of transcription and addition
C      of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
C      (2) determining a level or pattern of a molecule in a bovine cell or
C      tissue comprising: (a) incubating a marker nucleic acid (comprising any
C      of the 15112 nucleic acid sequences or its complement or fragment) with a
C      complementary nucleic acid molecule obtained from the bovine cell or
C      tissue, where hybridisation between the marker nucleic acid and the
C      complementary nucleic acid permits the detection of the molecule; and (b)
C      detecting the level or pattern of the complementary nucleic acid, where
C      the detection of the complementary nucleic acid is predictive of the
C      level or pattern of the molecule. The LMPD nucleic acid is used for
C      determining a level or pattern of a molecule in a bovine cell or tissue.
C      It is useful for genome mapping, gene identification and analysis, cattle
C      breeding, preparation of constructs for use in cattle gene expression, CR
C      for genetically improving cattle. The present sequence is one of the
C      15112 bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The
C      present sequence was not shown in the specification but was obtained in
C      electronic format from the USPTO web site:
C      seqdata.uspto.gov/sequence.html?DocID=20020137139
X      Sequence 403 BP; 90 A; 130 C; 114 G; 69 T; 0 U; 0 Other;
X      Alignment Scores:
X      pred. No.:          31.3           Length:         403
X      score:              69.50          Matches:         15
X      Percent Similarity: 48.48%          Conservative:    1
X      Best Local Similarity: 45.45%        Mismatches:     10
X      Query Match:       41.12%          Indels:         7
X      DB:                7               Gaps:           1
X      JS-10-010-050A-2 COPY_1_30 (1-30) x ABX40924 (1-403)
Y      4   GYALAGIYALALAAIGSYARGAIsERTTCys-----Tip 16
Y      283 GGAGCAGGCGACATCATGTCTCTGTGCATSTGCCGACTGGTGCGCTTTTCATGG 224
Y      17 AlAlenAlaleuTeuTrpleuAlaValAlProGlyTTP 29
Y      223 GCTCTGGCCCTGCATGGCAGCGCCGCCGATCAGCATGG 185
X      RESULT 15
X      ABX46110/c
X      ID      ABX46110 standard; cDNA; 404 BP.
X      AC      ABX46110;
X      CT      21-FEB-2003 (first entry)
X      Bovine EST associated with lactation/muscle/fat deposition #11275.

```

```

XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KM gene analysis; cattle breeding.
XX
XX Bos Taurus.
XX
XX US2002137139-A1.
XX
XX PD
XX 26-SEP-2002.
XX
XX PF 24-SEP-2001; 2001US-00960352.
XX
XX PR 12-JAN-1999; 99US-0115707P.
XX PR 11-JAN-2000; 2000US-00480902.
XX
XX PA (BYATT/) BYATT J C.
XX PA (MATH/) MATHIALAGAN N.
XX PA (TAON/) TAO N.
XX PA (WARRE/) WARREN W C.
XX
XX PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX DR MPI; 2003-110599/10.
XX
XX PT New nucleic acid associated with lactation, and muscle and fat
XX deposition, useful for genome mapping, gene identification and analysis,
XX cattle breeding, or for genetically improving cattle.
XX PS
XX Claim 2; SEQ ID NO 11275; 245pp; English.
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived from
CC cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC appearing as ABX34836-ABX49947, or complements of them. Also included are
CC : (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC acid linked to a promoter and a 3' non-translated sequence that
CC functions in the cell to cause termination of transcription and addition
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC (2) determining a level or pattern of a molecule in a bovine cell or
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC of the 15112 nucleic acid sequences or its complement or fragment) with a
CC complementary nucleic acid molecule obtained from the bovine cell or
CC tissue, where hybridisation between the marker nucleic acid and the
CC complementary nucleic acid permits the detection of the molecule; and (b)
CC detecting the level or pattern of the complementary nucleic acid, where
CC the detection of the complementary nucleic acid is predictive of the
CC level or pattern of the molecule. The LMFD nucleic acid is used for
CC determining a level or pattern of a molecule in a bovine cell or tissue.
CC It is useful for genome mapping, gene identification and analysis, cattle
CC breeding, preparation of constructs for use in cattle gene expression, or
CC for genetically improving cattle. The present sequence is one of the
CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC present sequence was not shown in the specification but was obtained in
CC electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
XX SQ Sequence 404 BP; 91 A; 131 C; 115 G; 67 T; 0 U; 0 Other;

```

Qy 17 AlaLeuAlaLeuLeuTrpLeuAlaValValProGlyTrp 29
|||
Db 244 GCTCTGGGCTTGCCTGCAATGGGCAAGCGCCGCATCAGCATGG 206
|||

Search completed: April 24, 2004, 23:26:34
Job time : 57.375 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 23:02:32 ; Search time 355.406 Seconds
(without alignments)
2352.639 Million cell updates/sec

Title: US-10-010-050a-2_COPY_1_28
Perfect score: 154
Sequence: 1 MRRGAGAGGRASWCALALLMLAVPG 28

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/comp1/uspto.spool/US10010050/runat_22042004_113204_27556/app.query.fasta_1.1372
-DB=EST -QPM=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0 -LONGLIST=45
-UNITS-bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human0.cdt -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFM=pcio -NORM=ext -HEAPSIE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US10010050 @CEN 1 1 6283 @runat_22042004_113204_27556 -NCPU=6 -ICPU=3
-MMP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLIST
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_pln:*
20: em_gss_vtl:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pio:*
25: em_gss_rtd:*
26: em_gss_plg:*
27: em_gss_vtl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	154	100.0	565	9	AU280382	AU280382 AU280382
2	154	100.0	627	12	BM849262	BM849262 K-EST0129
3	154	100.0	623	13	BQ417448	BQ417448 iK38b05.Y
4	154	100.0	676	12	BG820113	BG820113 602782325
5	154	100.0	787	12	B1223533	B1223533 602941923
6	154	100.0	847	13	BQ179074	BQ179074 AGENCOURT
7	154	100.0	869	12	BI090566	BI090566 602855673
8	154	100.0	923	13	BX31615	BX31615 BX31615
9	154	100.0	1098	10	BP982158	BP982158 602308976
10	154	100.0	1201	9	AL546472	AL546472 AL546472
11	154	100.0	1222	13	BQ897986	BQ897986 AGENCOURT
12	146	94.8	924	14	CA488543	CA488543 AGENCOURT
13	141	91.6	1077	13	EX342662	EX342662 BX342662
14	127	82.5	589	10	BP977311	BP977311 602146485
15	124	80.5	555	14	CA397574	CA397574 cs92h10.Y
16	104	67.5	273	9	AA300650	AA300650 EST13763
17	73	47.4	383	14	CB808908	CB808908 AMGNNUC.M
18	73	47.4	424	14	CB795968	CB795968 AMGNNUC.M
19	73	47.4	427	14	CB548129	CB548129 AMGNNUC.M
20	72	46.8	1355	10	BE880806	BE880806 601493233
21	71	46.1	569	29	CE827309	CE827309 tigr-gss-
22	71	46.1	632	13	BX844586	BX844586 BX844586
23	69.5	45.1	1258	13	BQ945568	BQ945568 AGENCOURT
24	69	44.8	367	13	BY014529	BY014529 BY014529
25	69	44.8	377	13	BY015054	BY015054 BY015054
26	69	44.8	395	13	BY027225	BY027225 BY027225
27	69	44.8	421	13	BY247307	BY247307 BY247307
28	69	44.8	451	10	BB639395	BB639395 BB639395
29	69	44.8	665	13	BY721389	BY721389 BY721389
30	69	44.8	686	14	CB651802	CB651802 OSUNEB160
31	69	44.8	703	14	CF876342	CF876342 trico41xb
32	69	44.8	722	14	CB660119	CB660119 OSUNED01T
33	69	44.8	771	14	CB905022	CB905022 trico41xb
34	69	44.8	796	12	B1182240	B1182240 UNL-P-FN-
35	68.5	44.5	571	12	BM171756	BM171756 imageqc.5
36	68.5	44.5	642	10	BF900979	BF900979 IL0-MT030
37	68.5	44.5	977	10	BF032072	BF032072 601559649
38	68	44.2	743	29	CNS01LHE	AL149635 Anopheles
39	68	44.2	773	28	BZ733787	BZ733787 OGDSS6TM
40	68	44.2	827	29	CNS01RBN	AL157095 Anopheles
41	67.5	43.8	411	13	B1227424	BY227424 BY227424
42	67.5	43.8	561	9	A1786906	A1786906 UJ32909.Y
43	67.5	43.8	587	10	BG077284	BG077284 H3013B06-
44	67.5	43.8	789	29	CG276043	CG276043 OGWF52TV
45	67.5	43.8	825	12	BG536228	BG536228 602565449

ALIGNMENTS

RESULT 1
AU280382
LOCUS AU280382
DEFINITION AU280382 NIBSE2 Homo sapiens cDNA clone NIBSE2001763 5', mRNA
SEQUENCE
ACCESSION AU280382
VERSION AU280382.1
KEYWORDS GI:28299609
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 565)

ORIGIN
 /clone_1lb="NIH_MGC_12"
 /note="Organ: cervix; Vector: PCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."

ALIGNMENT SCORES:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
3.55e-06	787	154.00	28	0	0	0
Score:	100.00%	100.00%	0	0	0	0
Percent Similarity:	100.00%	100.00%	0	0	0	0
Best Local Similarity:	100.00%	100.00%	0	0	0	0
Query Match:	100.00%	100.00%	0	0	0	0

US-10-010-050A-2_COPY_1_28 (1-28) x B1223533 (1-787)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
 DB 26 ATGCGGCGGGCGGGCGGGCGGCTCGGGGAGCGGCTTCGTGCTGCGGCTCGGGCTG 65

QY 21 LeuTrpLeuAlaValAlaProGly 28
 DB 86 CTTGGCTCGCGGTGTCCGAGC 109

RESULT 6 BQ719074 847 bp mRNA linear EST 16-JUL-2002
 LOCUS BQ719074
 DEFINITION AGENCOURT 8103813 lupski_sympathetic_trunk Homo sapiens cDNA clone

IMAGE:6191057 5', mRNA sequence.

ACCESSION BQ719074

VERSION BQ719074.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 847)

NIH-MGC http://mgs.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Dr. James R. Lupski

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: L1M13590 row: n column: 18

High quality sequence stop: 484.

Location/Qualifiers

1.847

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6191057"

/sex="male"

/tissue_type="sympathetic trunk"

/dev_stage="adult, 16 yr"

/lab_host="DH10B"

/clone_1lb="lupski_sympathetic_trunk"

/note="Vector: PCMV-SPORT6 (Life Technologies); Site 1:

NotI; Site 2: SalI; cDNA made by oligo-dT priming.

Directionally cloned using the following adaptors:

5'-TCGACCAAGGCTCCG-3' and

5'-GACTAGTTTATGATCGGAGCGGCCCTT(15)-3'. Size selected >

1 kb for average insert length 1.9 kb. This is a primary

library, non-amplified. Library constructed by Life

Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor

ORIGIN

ALIGNMENT SCORES:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
3.9e-06	847	154.00	28	0	0	0
Score:	100.00%	100.00%	0	0	0	0
Percent Similarity:	100.00%	100.00%	0	0	0	0
Best Local Similarity:	100.00%	100.00%	0	0	0	0
Query Match:	100.00%	100.00%	0	0	0	0

US-10-010-050A-2_COPY_1_28 (1-28) x BQ719074 (1-847)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
 DB 80 ATGCGGCGGGCGGGCGGGCGGCTCGGGGAGCGGCTTCGTGCTGCGGCTCGGGCTG 139

QY 21 LeuTrpLeuAlaValAlaProGly 28
 DB 140 CTTGGCTCGCGGTGTCCGAGC 163

RESULT 7 B1090566 869 bp mRNA linear EST 20-JUN-2001
 LOCUS B1090566
 DEFINITION 602853673F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996891 5',

mRNA sequence.

ACCESSION B1090566

VERSION B1090566.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 869)

NIH-MGC http://mgs.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: L1M11023 row: a column: 20

High quality sequence stop: 843.

Location/Qualifiers

1.869

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4996891"

/cell_line="MGC36"

/lab_host="DH10B"

/clone_1lb="NIH_MGC_10"

/note="Organ: Cervix; Vector: PCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.5 kb. Library prepared by Life

Technologies."

ORIGIN

ALIGNMENT SCORES:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4.03e-06	869	154.00	28	0	0	0
Score:	100.00%	100.00%	0	0	0	0
Percent Similarity:	100.00%	100.00%	0	0	0	0
Best Local Similarity:	100.00%	100.00%	0	0	0	0
Query Match:	100.00%	100.00%	0	0	0	0

US-10-010-050A-2_COPY_1_28 (1-28) x B1090566 (1-869)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20

```

b 52 ATGCGCGGGGCGCGGCGCGGCTCTCTGCTGCTGGGCGCTTGGGCGCTG 11:
|=====|
y 21 LeuTripleuAlaValAlaProGly 28
|=====|
b 112 CTTTGCGCTCGCGGTGTTCGGGCG 135
|=====|

RESULT 8
BX31615 923 bp mRNA linear EST 01-MAY-2003
LOCUS BX31615 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DB009YC01 5-PRIME, mRNA sequence.
ACCESSION BX31615 GI:30310073
VERSION BX31615.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 923)
Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f. For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DB009AB01Q1&cluster=7238.f. Contact :
Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DB009AB01Q1.
Location/Qualifiers
1. 923
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB009YC01"
/tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
/clone_1b="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
/Note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Alignment Scores:
Pred. No.: 4,35e-06 Length: 923
Score: 154.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

JS-10-010-050A-2_COPY_1_28 (1-28) x BX31615 (1-923)
2y 1 MetcArgArgGlyAlaGlyAlaAlaArgGlyAArgAlaSerTrpCySTrpAlaLeuAlaLeu 20
|=====|
yb 118 ATGCGCGGGGCGCGGCGGCTCTCTGCTGCTGGGCGCTTGGGCGCTG 117
|=====|
2y 21 LeuTripleuAlaValAlaProGly 28
|=====|
yb 178 CTTTGCGCTCGCGGTGTTCGGGCG 201
|=====|

RESULT 9
BF982158 1098 bp mRNA linear EST 23-JAN-2001
LOCUS BF982158 602308976P1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4400298 5',
DEFINITION mRNA sequence.
ACCESSION BF982158

```

```

VERSION BF982158.1 GI:12384970
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1098)
Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

ORIGIN
Alignment Scores:
Pred. No.: 5,44e-06 Length: 1098
Score: 154.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-010-050A-2_COPY_1_28 (1-28) x BF982158 (1-1098)
Qy 1 MetcArgArgGlyAlaGlyAlaAlaArgGlyAArgAlaSerTrpCySTrpAlaLeuAlaLeu 20
|=====|
Db 35 ATGCGCGGGGCGCGGCGGCTCTCTGCTGCTGGGCGCTTGGGCGCTG 94
|=====|
Qy 21 LeuTripleuAlaValAlaProGly 28
|=====|
Db 95 CTTTGCGCTCGCGGTGTTCGGGCG 118
|=====|

RESULT 10
AL546472 1201 bp mRNA linear EST 31-MAY-2003
LOCUS AL546472 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1030YJ01 5-PRIME, mRNA sequence.
ACCESSION AL546472
VERSION AL546472.2 GI:31268306
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1201)
Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:12879620.
Contact: Genoscope

```


hTERT-HMEL, LNCaP"
/lab host="EMD10B"
/clone lib="MAPCL"
/note="Vector: PCMV-SPORE6; Site 1: EcorV, Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Egland, James U. Vincent, Robert Strassberg,
Bungkok Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ALIGNMENT SCORES:
red. No.: 2.93e-05 Length: 924
core: 146.00 Matches: 27
Percent Similarity: 96.43% Conservative: 0
Best Local Similarity: 96.43% Mismatches: 1
Query Match: 94.81% Indels: 0
Gaps: 0

S-10-010-050A-2_COPY_1_28 (1-28) x CA488543 (1-924)

Y 1 MetAAGAGG1YALAG1YALAAGG1YARGA1ASerTPCySTRPAlaLeuAlaLeu 20
16 ATGTGGCGGGGGCGGGCGGGCGGCTTCTGAGTGGGCGCCCTGGCGCTG 75

Y 21 LeuTPLeuAlaValAlaProGly 28
76 CTTTGGCTCGCGGTGTTCCGGGC 99

RESULT 13
BX342662 1077 bp mRNA linear EST 02-MAY-2003
BX342662 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL004YRP03 5-PRIME, mRNA sequence.
ACCESSION BX342662
VERSION BX342662.1 GI:30338103
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1077)
Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f. For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DL004CH02QPI&cluster=7238.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DL004CH02QPI.

FEATURES
source
1. 1077
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL004YRP03"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and Ecor V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
ALIGNMENT SCORES:
Pred. No.: 0.000117 Length: 1077
Score: 141.00 Matches: 28
Percent Similarity: 96.55% Conservative: 0
Best Local Similarity: 96.55% Mismatches: 0
Query Match: 91.56% Indels: 1
Gaps: 0

US-10-010-050A-2_COPY_1_28 (1-28) x BX342662 (1-1077)

QY 1 MetAAGAGG1YALAG1YALAAGG1YARGA1ASerTPCySTRPAlaLeuAlaLeu 20
Db 115 ATGCGCGGGGGCGGGCGGGCGGCTTCTGAGTGGGCGCCCTGGCGCT 174

QY 20 ULeuTPLeuAlaValAlaProGly 28
Db 175 GCTTGGCTCGCGGTGTTCCGGGC 199

RESULT 14
BF977311 589 bp mRNA linear EST 22-JAN-2001
LOCUS BF977311
DEFINITION 602146485P1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4305501 5',
mRNA sequence.
ACCESSION BF977311 GI:12344526
VERSION BF977311.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 589)
NIH-MGC http://mhc.nci.nih.gov/.
TITLE NIH-MGC
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LNCMI172 row: 1 column: 22
High quality sequence scop: 586.

FEATURES
source
1. 589
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4305501"
/issue_type="melanotic melanoma, high MDR"
/lab host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_62"
/note="Organ: Skin. Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgctcgcc); Site_2: SfiI (ggccatcaggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGCGCGCATG-dT(30)BN-3'
(where B = A, C or G and N = A, C, G or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

ORIGIN
ALIGNMENT SCORES:
Pred. No.: 0.00152 Length: 589
Score: 127.00 Matches: 27
Percent Similarity: 96.43% Conservative: 0

Best Local Similarity: 96.43% Mismatches: 1
 Query Match: 82.47% Indels: 1
 DB: 10 Gaps: 0

US-10-010-050A-2_COPY_1_28 (1-28) x BP977311 (1-589)

Qy 1 MetArGArGGLyAlaGLyAlaAlaArgGLyArGAlaSeRTpCySTrPAlaLeuAlaLeu 20
 Db 161 ATGGCGCGGGCGCGGGCGCGGT CGGGAGCGCGTCTGCTGCTGCGCCCTGGCGCTG 219
 Qy 21 LeuTrPLeuAlaValAlaProGLy 28
 Db 220 CTTTGCTCGCGGTGCTCCGGGC 243

RESULT 15
 CA397574 555 bp mRNA linear EST 06-NOV-2002
 LOCUS cs92h10.y2 Human Retinal pigment epithelium/choroid cDNA
 DEFINITION (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs92h10
 5', mRNA sequence.

ACCESSION CA397574
 VERSION CA397574.1 GI:24735005
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 555)
 Wistow,G., Bernstein,S.I., Wyatt,M.K., Farris,R.N., Behal,A.,
 Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
 Expressed sequence tag analysis of human RPE/choroid for the
 NEIBank Project: Over 6000 non-redundant transcripts, novel genes
 and splice variants

Mol. Vis. 8 (4), 205-220 (2002)
 22103460
 12107410

JOURNAL MEDLINE PUBMED
 COMMENT Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 92 row: h column: 10
 Seq primer: M13RPI reverse primer (ABI).
 Location/Qualifiers

FEATURES
 source
 1..555
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="cs92h10"
 /tissue_type="RPE/choroid"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Human Retinal pigment epithelium/choroid cDNA
 (Un-normalized, unamplified): cs"
 /note="Organ: Eye; Vector: PCWVSPO6; Two different donor
 eyes (75-80 years old) yielded approximately 600 mg of
 dissected RPE/choroid tissue. This in turn yielded 340 ug
 of total RNA and 7 ug of mRNA. A directionally cloned cDNA
 library in the PCWVSPO6 vector was constructed at Life
 Technologies (Rockville, MD; now part of Invitrogen Corp),
 essentially following the protocols of the Superscript
 Plasmid System (Invitrogen Corp
 <http://www.invitrogen.com/>). The library code
 designation was cs. For this library, cDNA inserts were
 cloned into the NotI/Mlu sites of the vector. EST
 analysis was performed on the unamplified library at the
 NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:
 Pred. No.:

0.00288

Length:

555

Score: 124.00 Matches: 23
 Percent Similarity: 82.14% Conservative: 0
 Best Local Similarity: 82.14% Mismatches: 5
 Query Match: 80.52% Indels: 0
 DB: 14 Gaps: 0

US-10-010-050A-2_COPY_1_28 (1-28) x CA397574 (1-555)

Qy 1 MetArGArGGLyAlaGLyAlaAlaArgGLyArGAlaSeRTpCySTrPAlaLeuAlaLeu 20
 Db 40 ATGGCGCGGGCGCGGGCGCGGT CGGGAGCGCGTCTGCTGCTGCGCCCTGGCGCTG 99
 Qy 21 LeuTrPLeuAlaValAlaProGLy 28
 Db 100 CTTTGCTCGCGGTGCTCCGGGC 123

Search completed: April 25, 2004, 05:46:19
 Job time : 361.406 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 23:04:07 ; Search time 55.8056 Seconds

(without alignments)
2262.168 Million cell updates/sec

Title: US-10-010-050a-2_COPY_1_28
Percent score: 154
Sequence: 1 MRGAGAGRGRASWCALALLWLVVPG 28

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10010050/rnat_22042004_113205_27584/app.query.fasta_1.1372
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rmph -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10010050 @cgn_1_1_776 @rnat_22042004_113205_27584
-NCPU=6 -ICPU=3 -NO MAP -LARGOUEVRY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA:*

1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/2/pubpna/PCF_NEW_PUB.seq:*
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/2/pubpna/PCFUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	154	100.0	1486	9	US-09-122-383-1	Sequence 1, Appl
2	154	100.0	1486	14	US-10-010-050A-1	Sequence 1, Appl
3	154	100.0	1751	13	US-10-653-595-26	Sequence 26, Appl
4	154	100.0	1751	13	US-09-397-945-26	Sequence 26, Appl
5	106	68.8	1038	9	US-09-122-383-13	Sequence 13, Appl
6	106	68.8	1038	14	US-10-010-050A-13	Sequence 13, Appl
7	66	42.9	7155	15	US-10-329-079-14	Sequence 14, Appl
8	66	42.9	37350	15	US-10-329-079-6	Sequence 14, Appl
9	64	41.6	580	13	US-10-027-632-241716	Sequence 241716, Appl
10	64	41.6	580	16	US-10-027-632-241716	Sequence 241716, Appl
11	64	41.6	1071	13	US-10-282-122A-15190	Sequence 15190, A
12	63.5	41.2	276276	13	US-10-087-192-754	Sequence 754, App
13	63	40.9	772	15	US-10-029-517-106	Sequence 106, App
14	63	40.9	1257	13	US-10-425-114-5227	Sequence 5227, App
15	63	40.9	322101	13	US-10-354-247-1	Sequence 1, Appl
16	63	40.9	322101	15	US-10-060-902-1	Sequence 1, Appl
17	62.5	40.6	256	9	US-09-960-352-12900	Sequence 12900, A
18	62.5	40.6	1872	10	US-09-764-891-7060	Sequence 7060, App
19	62.5	40.6	1862	13	US-10-411-037-65	Sequence 65, Appl
20	62.5	40.6	1962	13	US-10-411-037-65	Sequence 65, Appl
21	62.5	40.6	2155	9	US-09-962-436-229	Sequence 299, App
22	62.5	40.6	2155	9	US-09-962-436-229	Sequence 2394, App
23	62.5	40.6	6200	9	US-09-993-038-1	Sequence 1, Appl
24	62.5	40.6	6200	9	US-09-993-038-1	Sequence 1, Appl
25	62.5	40.6	6200	15	US-10-206-443-1	Sequence 1, Appl
26	62.5	40.6	7648	17	US-10-176-066-1	Sequence 1, Appl
27	62.5	40.6	169998	16	US-10-380-931-24	Sequence 24, Appl
28	62.5	40.6	197496	9	US-09-877-177-10	Sequence 10, Appl
29	62.5	40.6	197496	13	US-10-426-836-10	Sequence 10, Appl
30	62	40.3	566	10	US-09-930-213-116	Sequence 116, App
31	62	40.3	2024	16	US-10-264-237-1269	Sequence 1269, App
32	61.5	39.9	404	9	US-09-960-352-6089	Sequence 6089, App
33	61.5	39.9	404	9	US-09-960-352-11275	Sequence 11275, A
34	61.5	39.9	414	9	US-09-960-352-14095	Sequence 14095, A
35	61.5	39.9	988	13	US-10-425-114-30764	Sequence 30764, A
36	61	39.6	558	13	US-10-027-632-67820	Sequence 67820, A
37	61	39.6	558	13	US-10-027-632-67820	Sequence 67820, A
38	61	39.6	558	13	US-10-027-632-67822	Sequence 67822, A
39	61	39.6	558	16	US-10-027-632-67820	Sequence 67820, A
40	61	39.6	558	16	US-10-027-632-67820	Sequence 67821, A
41	61	39.6	558	16	US-10-027-632-67821	Sequence 67821, A
42	61	39.6	96595	12	US-09-997-722-262	Sequence 262, App
43	61	39.6	684973	9	US-09-263-959-1	Sequence 1, Appl
44	60.5	39.3	596	9	US-09-811-284-39	Sequence 39, Appl
45	60.5	39.3	2506	13	US-10-282-122A-20063	Sequence 20063, A

ALIGNMENTS

RESULT 1
US-09-122-383-1
Sequence 1, Application US/09122383A
Patent No. US2002042093A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra C.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: CHROMOSOME 13
CURRENT FILING DATE: 1997-07-24
EARLIER APPLICATION NUMBER: US/09/122,383A
EARLIER FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1486
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (47)...(1084)

US-09-122-383-1

Alignment Scores:

Pred. No.:	8.32e-11	Length:	1486
Score:	154.00	Matches:	28
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-010-050a-2_COPY_1_28 (1-28) x US-09-122-383-1 (1-1486)

QY 1 MetAARGGlyYAlAGlyYAlaAlaArgGlyYArgAlaSerTTPCySTrPAlaLeuAlaLeu 20

Db 47 ATGCGCGGGGGCGGGCGGGCGGCTCGGGACCGCTTCTGTGCTGGGGCGGCTGGCGCTG 106

QY 21 LeuTrPLeuAlaValaValaProGly 28

Db 107 CTTGGCTCGCGGTGGTTCGGGCG 130

RESULT 2

US-10-010-050a-1

Sequence 1, Application US/10010050A
Publication No. US20020173624A1

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Gilbertson, Debra G.

TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN

FILE REFERENCE: CHROMOSOME 13

CURRENT APPLICATION NUMBER: US/10/010,050A

CURRENT FILING DATE: 2002-03-26

PRIOR APPLICATION NUMBER: US 09/122,383

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: US 60/053,613

PRIOR FILING DATE: 1997-07-24

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 1486

TYPE: DNA

ORGANISM: Homo sapien

FEATURE:

NAME/KEY: CDS

LOCATION: (47)...(1084)

US-10-010-050a-1

Alignment Scores:

Pred. No.:	8.32e-11	Length:	1486
Score:	154.00	Matches:	28
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050a-2_COPY_1_28 (1-28) x US-10-010-050a-1 (1-1486)

QY 1 MetAARGGlyYAlAGlyYAlaAlaArgGlyYArgAlaSerTTPCySTrPAlaLeuAlaLeu 20

Db 47 ATGCGCGGGGGCGGGCGGGCGGCTCGGGACCGCTTCTGTGCTGGGGCGGCTGGCGCTG 106

QY 21 LeuTrPLeuAlaValaValaProGly 28

Db 107 CTTGGCTCGCGGTGGTTCGGGCG 130

RESULT 3

US-10-653-595-26

Sequence 26, Application US/10653595
Publication No. US20040048304A1

GENERAL INFORMATION:

APPLICANT: Ruben et. al.

TITLE OF INVENTION: 95 Human secreted proteins

FILE REFERENCE: P2027P1C1

CURRENT APPLICATION NUMBER: US/10/653,595

CURRENT FILING DATE: 2003-09-03

PRIOR APPLICATION NUMBER: US 09/397945

PRIOR FILING DATE: 1999-09-17

PRIOR APPLICATION NUMBER: PCT/US99/05804

PRIOR FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: 60/078,566

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,576

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,573

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,574

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,579

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/080,314

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/080,312

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/078,578

PRIOR FILING DATE: 1998-03-19

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 470

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 26

LENGTH: 1751

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (1520)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1557)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1689)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1729)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1735)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1741)

OTHER INFORMATION: n equals a,t,g, or c

US-10-653-595-26

Alignment Scores:

Pred. No.:	9.82e-11	Length:	1751
Score:	154.00	Matches:	28
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-10-010-050a-2_COPY_1_28 (1-28) x US-10-653-595-26 (1-1751)

QY 1 MetAARGGlyYAlAGlyYAlaAlaArgGlyYArgAlaSerTTPCySTrPAlaLeuAlaLeu 20

Db 49 ATGCGCGGGGGCGGGCGGGCGGCTCGGGACCGCTTCTGTGCTGGGGCGGCTGGCGCTG 108

QY 21 LeuTrPLeuAlaValaValaProGly 28

Db 109 CTTGGCTCGCGGTGGTTCGGGCG 132


```

-10-010-050A-2_COPY_1_28 (1-28) x US-10-027-632-241716 (1-580)
      13
      Gaps: 0
      4 G|y|a|l|a|g|a|a|a|a|r|g|g|y|A|r|g|a|l|a|s|e|r|T|p|C|y|e|r|T|p|a|l|e|u|a|l|e|u|T|p|l|e|u
        |||||:::|||||
      22 GGTCAATGCCAGTCGCAAGAAGTGGGTTTCCATCGTGTGGGCAAGCTCCAACTCTGTGGCTT 81
      24 A|a|v|a|l|a|l 26
          |||
      82 TGCAGGGTA 90

SULT 10
-10-027-632-241716
Sequence 241716, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMERASE CHAIN REACTION (PCR) Assays for Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIORITY CLAIMING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 241716
LENGTH: 580
TYPE: DNA
ORGANISM: Human
-10-027-632-241716

Alignment Scores:
Seq. No.: 15 Length: 580
Score: 64.00 Matches: 11
Percent Similarity: 56.52% Conservative: 2
At Local Similarity: 47.83% Mismatches: 10
Any Match: 41.56% Indels: 0
Gaps: 0
: 16

-10-010-050A-2_COPY_1_28 (1-28) x US-10-027-632-241716 (1-580)
      4 G|y|a|l|a|g|a|a|a|a|r|g|g|y|A|r|g|a|l|a|s|e|r|T|p|C|y|e|r|T|p|a|l|e|u|a|l|e|u|T|p|l|e|u
        |||||:::|||||
      22 GGTCAATGCCAGTCGCAAGAAGTGGGTTTCCATCGTGTGGGCAAGCTCCAACTCTGTGGCTT 81
      24 A|a|v|a|l|a|l 26
          |||
      82 TGCAGGGTA 90

SULT 11
-10-282-122A-15190
Sequence 15190, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeelbeck, Robert
APPLICANT: Ohlsen, Karl
```

```

APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA-034A
CURRENT FILING DATE: 2003-02-20
PRIORITY APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15190
LENGTH: 1071
TYPE: DNA
ORGANISM: Bordetella pertussis
US-10-282-122A-15190

Alignment Scores:
Pred. No.:      27.9      Length:      1071
Score:          64.00     Matches:      11
Percent Similarity: 60.87% Conservative:   3
Best Local Similarity: 47.83% Mismatches:    9
Query Match:      41.56% Indels:         0
DB:               13      Gaps:           0
US-10-010-050A-2 COPY_1_28 (1-28) x US-10-282-122A-15190 (1-1071)
QY      2 ARGARGGLVAGLIALALaIaRgslYaRgaIsaeRTpCYSTPaLaLeuaIeUeu 21
Db      251 AGGCCGCGGGCAGGCTGCCTATCCGACCGACGTCGTCTCTGGCATGGCGCGTATG 310
        ::::::::::::::::::::
QY      22 TTPLeuaIA 24
        ::::::::::::::
Db      311 TGCGTTTCC 319

RESULT 12
US-10-087-192-754/C
Sequence 754, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
TITILE OF INVENTION: CANCER
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/799,586

```

;; PRIOR FILING DATE: 2001-03-02
;; NUMBER OF SEQ ID NOS: 2059
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 754
;; LENGTH: 276276
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-087-192-754

Alignment Scores:
Pred. No.: 8.84e+03
Score: 63.50
Length: 276276
Matches: 13
Percent Similarity: 52.00%
Conservative: 0
Best Local Similarity: 52.00%
Mismatch: 11
Query Match: 41.23%
Indels: 1
Gaps: 1

US-10-010-050a-2_COPY_1_28 (1-28) x US-10-087-192-754 (1-276276)

Qy 4 GlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeuLeu---Trp 23
Db 221348 GGAGCCAGCACCACCCGGGCTTCTTCAGCTGCTGGGCTGGGGCTGGCCCTCG 221289
Qy 23 LeuAlaValAlaPro 27
Db 221288 CTGTACTGGGTGCG 221274

RESULT 13

US-10-029-517-106
; Sequence 106, Application US/10029517
; Publication No. US20030148969A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 106
; LENGTH: 772
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-10-029-517-106

Alignment Scores:
Pred. No.: 27
Score: 63.00
Length: 772
Matches: 13
Percent Similarity: 60.00%
Conservative: 2
Best Local Similarity: 52.00%
Mismatch: 8
Query Match: 40.91%
Indels: 2
Gaps: 1

S-10-010-050a-2_COPY_1_28 (1-28) x US-10-029-517-106 (1-772)

Y 4 GlyAlaGlyAlaAlaArg-----GlyArgAlaSerTrpCysTrpAlaLeuAlaLeuLeu 21
b 551 GGAGCTGGGGTGGCCAGGCTGGGSCATGCGGCTCTGCTGGGCTGCTGCTGCTGCT 61C
Y 22 TrpLeuAlaValAla 26
b 611 GCGCTGGCCATTGTC 625

35UTR 14

3-10-425-114-5227/c
Sequence 5227, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E

;; APPLICANT: Tabaka, Jack E
;; APPLICANT: Cao, Yongwei
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;; FILE REFERENCE: 38-21(53313)B
;; CURRENT APPLICATION NUMBER: US/10/425,114
;; NUMBER OF SEQ ID NOS: 2003-04-28
;; SEQ ID NO 5227
;; LENGTH: 1257
;; TYPE: DNA
;; ORGANISM: Zea mays
;; FEATURE:
; OTHER INFORMATION: Clone ID: 700448276_FLI
US-10-425-114-5227

Alignment Scores:
Pred. No.: 44.1
Score: 63.00
Length: 1257
Matches: 10
Percent Similarity: 66.67%
Conservative: 0
Best Local Similarity: 66.67%
Mismatch: 5
Query Match: 40.91%
Indels: 0
Gaps: 0

US-10-010-050a-2_COPY_1_28 (1-28) x US-10-425-114-5227 (1-1257)

Qy 2 ArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrp 16
Db 881 GGAGCAGCAGCAGAGCGCGGCGAGCGCGCTTCTTGGGTGCG 837

RESULT 15

US-10-354-247-1/c
; Sequence 1, Application US/10354247
; Publication No. US20030224393A1
; GENERAL INFORMATION:
; APPLICANT: Gudmundsson Gudmundur
; TITLE OF INVENTION: GENE FOR PERIPHERAL ARTERIAL OCCLUSIVE
; FILE REFERENCE: 2345.2012-003
; CURRENT APPLICATION NUMBER: US/10/354,247
; NUMBER OF SEQ ID NOS: 36
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US 10/060,902
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 322101
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(322101)
; OTHER INFORMATION: n = A,T,C or G
US-10-354-247-1

Alignment Scores:
Pred. No.: 1.2e+04
Score: 63.00
Length: 322101
Matches: 11
Percent Similarity: 60.00%
Conservative: 1
Best Local Similarity: 55.00%
Mismatch: 8
Query Match: 40.91%
Indels: 0
Gaps: 0

US-10-010-050a-2_COPY_1_28 (1-28) x US-10-354-247-1 (1-322101)

Qy 4 GlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeuLeuTrpLeu 23
Db 166506 GGTGAGCTGACGACGAGAGGTAGTCCATGCTTGGGAGCTCCACTTGTGGCTT 166447

Search completed: April 25, 2004, 06:14:49
Job time : 95.8056 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 23:02:52 ; (without time 9.29444 Seconds)

1671.819 Million cell updates/sec

Title: US-10-010-050a-2_COPY_1_28
Perfect score: 154
Sequence: 1 MRGAGARGARASWCWALILMLAVPG 28

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h
-Q=/cg2_1/USPTO.spool/US10010050/runat_22042004_113205_27569/app_query.fasta_1.1372
-DB=Issued_Patents_NA -OPMT=Isatap -SUFFIX=ini -MINMATCH=0.1 -LOOPL=0
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10010050 @CGN 1.1 140 @runat_22042004_113205_27569 -NCPN=6 -ICPU=3
-NO_MMAP -JARBQOTRY -NGG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued Patents_NA.*
2: /cg2_6/prodata/2/ina/5A_COMB.seq.*
3: /cg2_6/prodata/2/ina/5B_COMB.seq.*
4: /cg2_6/prodata/2/ina/6A_COMB.seq.*
5: /cg2_6/prodata/2/ina/6B_COMB.seq.*
6: /cg2_6/prodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	62.5	40.6	2067	US-08-713-928B-8	Sequence 8, Appl1
C 2	62.5	40.6	2155	US-09-191-171-4	Sequence 4, Appl1
C 3	62.5	40.6	2155	US-09-385-707-4	Sequence 4, Appl1
C 4	62.5	40.6	2155	US-09-639-696C-4	Sequence 4, Appl1
C 5	62.5	40.6	4480	US-09-191-171-7	Sequence 7, Appl1
C 6	62.5	40.6	4480	US-09-385-707-7	Sequence 7, Appl1
C 7	62.5	40.6	6200	US-09-439-923-1	Sequence 1, Appl1
C 8	62.5	40.6	6200	US-09-711-202A-1	Sequence 1, Appl1
C 9	62.5	40.6	6200	US-09-711-202A-1	Sequence 1, Appl1
C 10	62.5	40.6	6238	US-09-639-696C-6	Sequence 6, Appl1
C 11	62.5	40.6	16998	US-09-676-610B-24	Sequence 24, Appl1
C 12	62.5	40.6	197496	US-09-877-177A-10	Sequence 10, Appl1

C 13	62	40.3	365	US-09-621-976-11459	Sequence 11459, A
C 14	61	39.6	2502	US-09-059-023-2	Sequence 2, Appl1
C 15	60.5	39.3	657	US-09-489-039A-2986	Sequence 2986, Ap
C 16	60.5	39.3	1452	US-09-489-039A-2735	Sequence 2735, Ap
C 17	60	39.0	894	US-09-489-039A-5631	Sequence 5631, Ap
C 18	60	39.0	1113	US-09-489-039A-5264	Sequence 5264, Ap
C 19	60	39.0	33529	US-09-144-085-3	Sequence 3, Appl1
C 20	59	38.3	1017	US-09-023-655-738	Sequence 738, Ap
C 21	59	38.3	1731	US-09-489-039A-688	Sequence 688, Ap
C 22	58	37.7	1371	US-09-489-039A-1161	Sequence 1161, Ap
C 23	58	37.7	2100	US-09-020-830-2	Sequence 2, Appl1
C 24	58	37.7	2100	US-09-020-222-2	Sequence 2, Appl1
C 25	58	37.7	111282	US-09-754-250-3	Sequence 3, Appl1
C 26	57.5	37.3	1037	US-09-016-434-1121	Sequence 1121, Ap
C 27	57.5	37.3	37030	US-08-311-731A-25	Sequence 25, Appl1
C 28	57	37.0	1263	US-09-252-991A-11039	Sequence 11039, A
C 29	57	37.0	1656	US-09-252-991A-11216	Sequence 11216, A
C 30	57	37.0	2799	US-08-446-794A-5	Sequence 5, Appl1
C 31	57	37.0	2799	US-08-750-007-4	Sequence 4, Appl1
C 32	56.5	36.7	53165	US-09-214-808-1	Sequence 1, Appl1
C 33	56	36.4	381	US-09-252-991A-11793	Sequence 11793, A
C 34	56	36.4	395	US-09-702-705-831	Sequence 831, Ap
C 35	56	36.4	395	US-09-736-457-831	Sequence 831, Ap
C 36	56	36.4	395	US-09-614-124B-831	Sequence 831, Ap
C 37	56	36.4	395	US-09-671-325-831	Sequence 831, Ap
C 38	56	36.4	421	US-09-641-638-506	Sequence 506, Ap
C 39	56	36.4	700	US-08-037-579A-4	Sequence 4, Appl1
C 40	56	36.4	700	US-08-601-184-4	Sequence 4, Appl1
C 41	56	36.4	900	US-09-589-927-1	Sequence 1, Appl1
C 42	56	36.4	900	US-09-589-927-5	Sequence 5, Appl1
C 43	56	36.4	900	US-09-589-927-7	Sequence 7, Appl1
C 44	56	36.4	900	US-09-589-927-9	Sequence 9, Appl1
C 45	56	36.4	900	US-09-589-927-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1
US-08-713-928B-8/C
Sequence 8, Application US/08713928B
Patent No. 5929304
GENERAL INFORMATION:
APPLICANT: RADIN, DAVID N.
APPLICANT: CRAMER, CAROLE L.
APPLICANT: OISHI, KAREN K.
APPLICANT: WEISSENBORN, DEBORAH L.
TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSER: Penite & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713, 928B
FILING DATE: 13-SEP-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/003, 737
FILING DATE: 14-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7956-0011-999
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2067 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-713-928B-8

Alignment Scores:
Pred. No.: 160 Length: 2067
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 2 Gaps: 1

US-10-010-050A-2_COPY_1_28 (1-28) x US-08-713-928B-8 (1-2067)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
Db 1081 CGAAGGGGTGGCGGTGCTAGCAGAGGCATTGCTGCTCAGAGCGGTAGGGGA 1022

QY 11 ArgAlaSerTyrCysTrrpAlaLeuAla 19
Db 1021 AGCGCGAGGTGTGTGTGGCCAGTAGCA 995

RESULT 2
US-09-191-171-4/c
Sequence 4, Application US/09191171
Patent No. 6149909
Patent No. 6149909 6143294
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
TITLE OF INVENTION: SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,171
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993

ATTORNEY/AGENT INFORMATION:
NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8978Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 89..2047
US-09-191-171-4

Alignment Scores:
Pred. No.: 167 Length: 2155
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 3 Gaps: 1

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-191-171-4 (1-2155)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
Db 1169 CGAAGGGGTGGCGGTGCTAGCAGAGGCATTGCTGCTCAGAGCGGTAGGGGA 1110

QY 11 ArgAlaSerTyrCysTrrpAlaLeuAla 19
Db 1109 AGCGCGAGGTGTGTGTGGCCAGTAGCA 1083

RESULT 3
US-09-385-707-4/c
Sequence 4, Application US/09385707
Patent No. 6238662
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
TITLE OF INVENTION: SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,707
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,104
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8978Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 89..2047
US-09-385-707-4

Alignment Scores:
Pred. No.: 167 Length: 2155
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
Gaps: 1
DB: 3

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-385-707-4 (1-2155)

Qy 2 ArgArgGlyAla-----GlyAlaAlaArggly 10
Db 1169 CGAAGGGGTGGCGGTGAGTCTCAGGAAGCATTTGCTGCTCAGAGCGGTAGGGGA 1110

Qy 11 ArgAlaSerTyrCysTyrAlaLeuAla 19
Db 1109 AGCGGAGGTGGTGTGGCCAGTAGCA 1083

RESULT 4
US-09-639-696C-4/c
Sequence 4, Application US/09639696C
Patent No. 6524835
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
ANSON, Donald S.
ORSBORN, Annette M.
NELSON, Paul V.
CLEMENTS, Peter R.
MORRIS, Charles P.
HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON PEABODY LLP
STREET: 990 Stewart Avenue
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/639,696C
FILING DATE: 16-AUG-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/191,171
FILING DATE: 13-NOV-1998
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995

APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pokalsky, Ann R.
REGISTRATION NUMBER: 34,697
REFERENCE/DOCKET NUMBER: 2249/304
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-832-7572
TELEFAX: 516-832-7555
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 89..2047
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-639-696C-4

Alignment Scores:
Pred. No.: 167 Length: 2155
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
Gaps: 1
DB: 4

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-639-696C-4 (1-2155)

Qy 2 ArgArgGlyAla-----GlyAlaAlaArggly 10
Db 1169 CGAAGGGGTGGCGGTGAGTCTCAGGAAGCATTTGCTGCTCAGAGCGGTAGGGGA 1110

Qy 11 ArgAlaSerTyrCysTyrAlaLeuAla 19
Db 1109 AGCGGAGGTGGTGTGGCCAGTAGCA 1083

RESULT 5
US-09-191-171-7/c
Sequence 7, Application US/09191171
Patent No. 6149909
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
ANSON, Donald S.
ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: MORRIS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/191,171
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-191-171-7

Alignment Scores:
Pred. No.: 375 Length: 4480
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 3 Gaps: 1

US-10-010-050a-2_copy_1_28 (1-28) x US-09-191-171-7 (1-4480)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
Db 2192 CGAAGGGGTCGGGTGAGTCAGGAGGCGATTGTCGTCACAGAGCGCGTAGGGGA 2133

QY 11 ArgAlaserTPrCystrPaAlaLeuAla 19
Db 2132 AGCGGAGGTGTGTGTGGCCAGTAGCA 2106

RESULT 6
US-09-385-707-7/c
Sequence 7, Application US/09385707
Patent No. 6238662
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John T.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
NUMBER OF SEQUENCES: 7
SEQUENCES ENCODING SAME
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,707
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,104
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-385-707-7

Alignment Scores:
Pred. No.: 375 Length: 4480
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 3 Gaps: 1

US-10-010-050a-2_copy_1_28 (1-28) x US-09-385-707-7 (1-4480)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
Db 2192 CGAAGGGGTCGGGTGAGTCAGGAGGCGATTGTCGTCACAGAGCGCGTAGGGGA 2133

QY 11 ArgAlaserTPrCystrPaAlaLeuAla 19
Db 2132 AGCGGAGGTGTGTGTGGCCAGTAGCA 2106

RESULT 7
US-09-439-923-1/c
Sequence 1, Application US/09439923
Patent No. 6426208
GENERAL INFORMATION:
APPLICANT: Emil D. Kakkis
APPLICANT: Becky Tanamachi
TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods for
Producing and Purifying the Same and Methods for
Treating Diseases Caused by Deficiencies Thereof
FILE REFERENCE: 08000051US00
CURRENT APPLICATION NUMBER: US/09/439,923
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 6200
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1558)...(3516)
US-09-439-923-1

Alignment Scores:
Pred. No.: 537 Length: 6200
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
```

```
test Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 4 Gaps: 1
US-10-010-050A-2_COPY_1_28 (1-28) x US-09-439-923-1 (1-6200)

2y 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
   |||||
   2638 CGAAGGGGTGCGGTGTAGCTCAGGAAGCATTTGTTGTCACAGAGCGCGTAGGGGA 2579
   |||||
   2578 AGGCGGAGGTGTTGTCGCGCAGTAGCA 2552

2y 11 ArgAlaSerTrpCysTrpAlaLeuAla 19
   |||||
   2578 AGGCGGAGGTGTTGTCGCGCAGTAGCA 2552

RESULT 8
US-09-711-202A-1/c
Sequence 1, Application US/09711202A
Patent No. 6569661
GENERAL INFORMATION:
APPLICANT: Emili D. Kakkis
TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods
TITLE OF INVENTION: for Producing and Purifying the Same and Methods for
TITLE OF INVENTION: Treating Diseases Caused by Deficiencies Thereof
FILE REFERENCE: 08000051US00
CURRENT APPLICATION NUMBER: US/09/711.202A
PRIOR FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: US/09/439,923
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 6200
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1558)...(3516)
US-09-711-202A-1

Alignment Scores:
Pred. No.: 537 Length: 6200
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 4 Gaps: 1

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-711-202A-1 (1-6200)

2y 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
   |||||
   2638 CGAAGGGGTGCGGTGTAGCTCAGGAAGCATTTGTTGTCACAGAGCGCGTAGGGGA 2579
   |||||
   2578 AGGCGGAGGTGTTGTCGCGCAGTAGCA 2552

2y 11 ArgAlaSerTrpCysTrpAlaLeuAla 19
   |||||
   2578 AGGCGGAGGTGTTGTCGCGCAGTAGCA 2552

RESULT 9
US-09-711-205A-1/c
Sequence 1, Application US/09711205A
Patent No. 6585971
GENERAL INFORMATION:
APPLICANT: Emili D. Kakkis
TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods
TITLE OF INVENTION: for Producing and Purifying the Same and Methods for
TITLE OF INVENTION: Treating Diseases Caused by Deficiencies Thereof
FILE REFERENCE: 08000051US00
CURRENT APPLICATION NUMBER: US/09/711.205A
PRIOR FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: US/09/439,923
PRIOR FILING DATE: 1999-11-12
```

```
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 6200
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1558)...(3516)
US-09-711-205A-1

Alignment Scores:
Pred. No.: 537 Length: 6200
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 4 Gaps: 1

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-711-205A-1 (1-6200)

Qy 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
   |||||
   2638 CGAAGGGGTGCGGTGTAGCTCAGGAAGCATTTGTTGTCACAGAGCGCGTAGGGGA 2579
   |||||
   2578 AGGCGGAGGTGTTGTCGCGCAGTAGCA 2552

Qy 11 ArgAlaSerTrpCysTrpAlaLeuAla 19
   |||||
   2578 AGGCGGAGGTGTTGTCGCGCAGTAGCA 2552

RESULT 10
US-09-639-696C-6/c
Sequence 6, Application US/09639696C
Patent No. 6524835
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
ANSON, Donald S.
ORSBORN, Annette M.
NELSON, Paul V.
CLEMENTS, Peter R.
MORRIS, Charles P.
HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: NIXON PEARBODY LLP
STREET: 990 Stewart Avenue
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/639,696C
FILING DATE: 16-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/191,171
FILING DATE: 13-NOV-1998
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pokalsky, Ann R.
```



```
DB      67539 ACCCTGCTCTCT---GCACCTGGA 67519
RESULT 12
US-09-877-177A-10/c
; Sequence 10, Application US/09877177A
; Patent No. 6582919
; GENERAL INFORMATION:
; APPLICANT: K. Danenberg
; TITLE OF INVENTION: Method of determining Epidermal Growth
; TITLE OF INVENTION: Factor Receptor and HER2-Neu Gene Expression
; TITLE OF INVENTION: and Correlation of Levels Theretof With Survival
; FILE REFERENCE: 11220/120
; CURRENT APPLICATION NUMBER: US/09/877,177A
; CURRENT FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 197496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-877-177A-10

Alignment Scores:
Pred. No.:      2.43e+04      Length:      197496
Score:          62.50         Matches:      13
Percent Similarity: 53.57%      Conservative: 2
Best Local Similarity: 46.43%      Mismatches:  12
Query Match:    40.58%         Indels:      1
DB:             4            Gaps:          1

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-877-177A-10 (1-197496)
QY      1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 10
DB      75599 ATGACACACGAGCGGCTCTACACAGCGCGTGACAGTGACCTGGGGTCTGACGCC 75540
QY      21 LeuTrpLeuAlaValAlaValProGly 28
DB      75539 ACCCTGCTCTCT---GCACCTGGA 75519

RESULT 13
US-09-621-976-11459/c
; Sequence 11459, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jober, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 11459
; LENGTH: 365
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-11459

Alignment Scores:
Pred. No.:      26.9         Length:      365
Score:          62.00         Matches:      12
Percent Similarity: 56.00%      Conservative: 2
Best Local Similarity: 48.00%      Mismatches:  11
Query Match:    40.26%         Indels:      0
DB:             4            Gaps:          0

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-621-976-11459 (1-365)
QY      2 ArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 21
```

```
DB      352 CGCATTTGGGTATATGACCAAGAGTAGTCCCATGGTCTTGGGACAGCTCGCCCC 293
QY      22 TrpLeuAlaValVal 26
DB      292 TGGCTCTGCGAGGTT 278

RESULT 14
US-09-069-023-2/c
; Sequence 2, Application US/09069023A
; Patent No. 6348573
; GENERAL INFORMATION:
; APPLICANT: Nunez, Gabriel
; APPLICANT: Inohara, Naohiro
; APPLICANT: Koseki, Takeyoshi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
; TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS
; FILE REFERENCE: UM-03333
; CURRENT APPLICATION NUMBER: US/09/069,023A
; CURRENT FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2502
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-069-023-2

Alignment Scores:
Pred. No.:      294         Length:      2502
Score:          61.00         Matches:      14
Percent Similarity: 58.06%      Conservative: 4
Best Local Similarity: 45.16%      Mismatches:  9
Query Match:    39.61%         Indels:      4
DB:             4            Gaps:          1

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-069-023-2 (1-2502)
QY      2 ArgArgGlyAlaGlyAlaAlaArg-----GlyArgAlaSerTrpCysTrpAla 17
DB      307 CCCCCGGCTCAGTAGCCGACAGGTGCGGAGATTGTGTGTAAGGAATGTGGCGATGGCG 248
QY      18 LeuAlaLeuLeuTrpLeuAlaValValProGly 28
DB      247 CTGCAGATGGCCTCCCGCTTCATGTCCTCCGGC 215

RESULT 15
US-09-489-039A-2986
; Sequence 2986, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2986
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2986

Alignment Scores:
Pred. No.:      76.8         Length:      657
Score:          60.50         Matches:      13
Percent Similarity: 55.56%      Conservative: 2
Best Local Similarity: 48.15%      Mismatches:  11
Query Match:    39.29%         Indels:      1
DB:             4            Gaps:          1
```

US-10-010-050A-2_COPY_1_28 (1-28) X US-09-489-039A-2986 (1-657)

Qy	3	ArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu---Leu	21
Db	224	CGTGGATTGAGGCCCGGTACGGGCGCGCGTCTGTGCTGTGCTGTGCGACCGGCTGT	283
Qy	22	TripleAlaValValProGly	28
Db	284	TCTATGGCATTGGCACCGGCG	304

Search completed: April 25, 2004, 05:51:38
 Job time : 46.2944 secs


```
PT in genetic manipulation.
XX
XX Example 11; SEQ ID NO 1884; 1380bp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence was used as the representative sequence
CC from a human clone which was used in homology searches to identify the
CC clone. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in CD-ROM format directly from
CC EPO
XX
SQ Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;
XX
Alignment Scores:
Pred. No.: 4.41e-09 Length: 697
Score: 154.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
XX
US-10-010-050A-2_COPY_1_28 (1-28) x AAK93424 (1-697)
XX
QY 1 MeAARGG1YAlaG1YAlaAlaArgG1YArgAlaSerTrpCyetrAlaLeuAlaLeu 20
DB 52 ATCGCGCGGGCGCGCGCGCGGCTCGGAGACGCGCTTCTGTGCTGGGCGCTGGGCGCTG 111
XX
QY 21 LeuTrpLeuAlaValValProGly 28
DB 112 CTTTGCTCGCGGCTGCTTCCGGGC 135
XX
RESULT 2
AAK92158
ID AAK92158 standard; cDNA; 697 BP.
XX
AC AAK92158;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 618.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
XX
PR 11-JAN-2000; 2000JP-00118774.
XX
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Itoai T, Hayashi K, Ishii S, Kawai Y;
XX
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuka T, Koga H;
XX
DR WPI: 2001-52425/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
XX
PT in genetic manipulation.
XX
PS Claim 2; SEQ ID NO 618; 1380bp + Sequence Listing; English.
```

```
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been isolated
XX and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
XX been determined. Primers for synthesizing the full length cDNA are useful
XX for clarifying the function of the protein encoded by the cDNA. The full
XX length clones were obtained by construction of full length enriched cDNA
XX libraries that were synthesised by the oligo-capping method. The primers
XX enable the production of the full length cDNA easily without any special
XX methods. The present sequence is the nucleotide sequence of the 5'-end of
XX a cDNA provided in the invention. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in CD-
XX ROM format directly from EPO
XX
SQ Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;
XX
Alignment Scores:
Pred. No.: 4.41e-09 Length: 697
Score: 154.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
XX
US-10-010-050A-2_COPY_1_28 (1-28) x AAK92158 (1-697)
XX
QY 1 MeAARGG1YAlaG1YAlaAlaArgG1YArgAlaSerTrpCyetrAlaLeuAlaLeu 20
DB 52 ATCGCGCGGGCGCGCGCGCGGCTCGGAGACGCGCTTCTGTGCTGGGCGCTGGGCGCTG 111
XX
QY 21 LeuTrpLeuAlaValValProGly 28
DB 112 CTTTGCTCGCGGCTGCTTCCGGGC 135
XX
RESULT 3
AAK02855
ID AAK02855 standard; DNA; 1486 BP.
XX
AC AAK02855;
XX
DT 14-MAY-1999 (first entry)
XX
DE Human zsig46 DNA.
XX
KW Secreted protein; zsig46; human; chromosome 13; thyroid; disease;
XX
KW hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;
XX
KW Hirschsprung's disease; neuronal ceroid-lipofusiosis; Wilson disease;
XX
KW Reiger syndrome; immunosassay; detection; anti-idiotypic antibody;
XX
KW therapy; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN
XX
FH Key Location/Qualifiers
XX
FT CDS 47..1087
XX
FT /*tag= a
XX
FT /product= "zsig46"
XX
XX
XX W09905275-A1.
XX
XX
XX 04-FEB-1999.
XX
XX
XX 24-JUL-1998; 98WO-US015431.
XX
XX
XX 24-JUL-1997; 97US-0053613P.
XX
XX
XX (ZYMO ) ZYMOGENETICS INC.
XX
XX
XX Shepard PO, Gilbertson DG;
XX
XX
XX WPI: 1999-142930/12.
XX
XX P-PSDB; AAW92967.
XX
XX
XX New secreted polypeptide, zsig46, and its fragments, related fusion
```

PT proteins - used for diagnosis and treatment of thyroid disorders or
 PT diseases involving genes on chromosome 13.
 XX
 PS Claim 27, Page 88-90; 101pp; English.
 XX
 CC This invention describes the isolation of a novel human secreted protein,
 CC zs1946 encoded by a gene on chromosome 13 which is mainly expressed in
 CC the thyroid. This product can be used to study secretion of proteins from
 CC cells and also to treat or prevent deficient expression of zs1946, which
 CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
 CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
 CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's
 CC disease, neuronal ceroid-lipofusiosis, Wilson disease and Reiger
 CC syndrome). Antibodies and other binding proteins, are used as immunoassay
 CC reagents to detect zs1946 or cells expressing it, e.g. for assessing
 CC purification of zs1946, to screen expression libraries, for affinity
 CC purification of zs1946, to deliver toxins, radioisotopes etc. for
 CC therapeutic or diagnostic purposes. Agonists of the product can be used
 CC to promote growth, differentiation and proliferation of specific cell
 CC types, e.g. for treating (extra)thyroid diseases or as additive to cell
 CC cultures
 CC
 XX Sequence 1486 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 1.03e-08 Length: 1486
 Score: 154.00 Matches: 28
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-10-010-050A-2_COPY_1_28 (1-28) x AA02855 (1-1486)
 QY 1 MetArgArgGlyAlaGlyAlaAlaAArgGlyAArgAlaSerTProCyTTPAlaLeuAlaLeu 20
 Db 47 ATGCGCGCGGCGCGCGCGCGCGCTCGGGAGACCGCTTCGTGGTGGGCGCGCGCTG 106
 QY 21 LeuTTPLeuAlaAlaValProGly 28
 Db 107 CTTTGCGCTCGCGGTGCTCCGGGC 130
 DE
 RESULT 4
 AA224826
 ID AA224826 standard; DNA; 1751 BP.
 XX
 AC AA224826;
 XX
 DT 02-DEC-1999 (first entry)
 XX
 DE Human secreted protein gene 16 clone HM2AD77.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumor; neurodegenerative disorder; leukemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; aschma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 KW
 OS Homo sapiens.
 XX
 PN WO9947540-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 18-MAR-1999; 99WO-US005804.
 XX
 PR 19-MAR-1998; 98US-0078563P.
 PR 19-MAR-1998; 98US-0078566P.
 PR 19-MAR-1998; 98US-0078573P.

PR 19-MAR-1998; 98US-0078574P.
 PR 19-MAR-1998; 98US-0078576P.
 PR 19-MAR-1998; 98US-0078577P.
 PR 19-MAR-1998; 98US-0078578P.
 PR 19-MAR-1998; 98US-0078579P.
 PR 19-MAR-1998; 98US-0078581P.
 PR 01-APR-1998; 98US-0080312P.
 PR 01-APR-1998; 98US-0080313P.
 PR 01-APR-1998; 98US-0080314P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
 PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Tafleur DM, Olsen HS;
 PI Shi Y, Moore PA;
 XX
 DR MPI: 1999-562050/47.
 DR P-PSDB; AA01323.
 DR
 PT New isolated human genes, useful for diagnosis and treatment of e.g.
 PT cancers, neurological disorders, immune diseases, inflammation or blood
 PT disorders.
 PT
 XX
 PS Claim 1, Page 308; 484pp; English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a secreted
 CC human protein. The gene number, and the clone it is derived from, are
 CC detailed in the descriptor line. The gene can be used to generate fusion
 CC proteins by linking to the gene to a human immunoglobulin Fc portion
 CC (e.g. AA224802) for increasing the stability of the fused protein as
 CC compared to the human protein only. The invention relates to 95 novel
 CC genes and their fragments (nucleic acid sequences: AA224811-224907; amino
 CC acid sequences AA01308-Y41404) which are useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,
 CC pathological conditions can be diagnosed by determining the amount of the
 CC new polypeptides in a sample or by determining the presence of mutations
 CC in the new polynucleotides. Specific uses are described for each of the
 CC 95 polynucleotides, based on which tissues they are most highly expressed
 CC in (see AA224811 for described uses)
 CC
 XX
 SQ Sequence 1751 BP; 477 A; 355 C; 380 G; 523 T; 0 U; 16 Other;
 Alignment Scores:
 Pred. No.: 1.24e-08 Length: 1751
 Score: 154.00 Matches: 28
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-10-010-050A-2_COPY_1_28 (1-28) x AA224826 (1-1751)
 QY 1 MetArgArgGlyAlaGlyAlaAlaAArgGlyAArgAlaSerTProCyTTPAlaLeuAlaLeu 20
 Db 49 ATGCGCGCGGCGCGCGCGCGCGCTCGGGAGACCGCTTCGTGGTGGGCGCGCTG 108
 QY 21 LeuTTPLeuAlaAlaValProGly 28
 Db 109 CTTTGCGCTCGCGGTGCTCCGGGC 132
 DE
 RESULT 5
 AA224829
 ID AA224829 standard; cDNA; 2120 BP.
 XX
 AC AA224829;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human full-length cDNA, SEQ ID NO: 3977.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 KW
 OS Homo sapiens.


```
KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
KM SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
XX OS  
XX PN WO200218632-A2.  
XX PD 07-MAR-2002.  
XX PF 01-SEP-2001; 2001WO-EP010074.  
XX PR 01-SEP-2000; 2000DE-01043826.  
XX PR 05-SEP-2000; 2000DE-01044543.  
XX PA (EPIG-) EPIGENOMICS AG.  
XX PI Olek A, Piepenbrock C, Berlin K, Gnetig D;  
XX DR WPI; 2002-371829/40.  
XX PT Determining the degree of cytosine methylation in genomic DNA, useful for  
XX PT diagnosis and prognosis, comprises selective hybridization of amplicons  
XX PT from chemically treated DNA.  
XX PS Claim 12; 56pp + Sequence listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of  
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
XX CC genomic sample of DNA. The sample is treated chemically to convert  
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
XX CC DNA that contains the target C is amplified to form a labeled amplicon.  
XX CC The amplicon is hybridized to two classes, each with at least one member,  
XX CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
XX CC degree of hybridization to both classes is determined from the label on  
XX CC the amplicon. From the ratio of labels hybridized to the two classes of  
XX CC oligomers, the degree of methylation is calculated. The method is used:  
XX CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
XX CC and of a wide range of diseases, e.g. cancer, disorders of the central  
XX CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
XX CC particularly by detecting mutations or single nucleotide polymorphisms  
XX CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
XX CC investigating cell differentiation. The method allows the methylation  
XX CC status of many C residues to be determined simultaneously. ABQ13410-  
XX CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
XX CC for determining the degree of cytosine methylation described in the  
XX CC disclosure of the invention  
XX  
SQ Sequence 630 BP; 86 A; 70 C; 255 G; 219 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 3.67e-05 Length: 630  
Score: 120.00 Matches: 21  
Percent Similarity: 75.00% Conservative: 0  
Best Local Similarity: 75.00% Mismatches: 7  
Query Match: 77.92% Indels: 0  
DB: Gaps: 0  
  
US-10-010-050A-2_COPY_1_28 (1-28) x ABQ40652 (1-630)  
QY 1 MetAArgGlyAlaAlaArgGlyArgAlaSerTPCyGTTTAlaLeuAlaLeu 20  
DB 299 ATGCGCGGGGCGCGGGCGGCTTTTGGTGTGGGTTTGGGCTTG 358  
|||  
QY 21 LeuTTPLeuAlaValValProGly 28  
|||  
DB 359 TTTTGGTTGCGCGGTTCGCGGT 382  
|||  
RESULT 8  
ABQ13702  
ID ABQ13702 standard; DNA; 631 BP.  
XX
```

```
AC ABQ13702;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 293.  
DE  
XX  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
KM SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
XX OS  
XX PN WO200218632-A2.  
XX PD 07-MAR-2002.  
XX PF 01-SEP-2001; 2001WO-EP010074.  
XX PR 01-SEP-2000; 2000DE-01043826.  
XX PR 05-SEP-2000; 2000DE-01044543.  
XX PA (EPIG-) EPIGENOMICS AG.  
XX PI Olek A, Piepenbrock C, Berlin K, Gnetig D;  
XX DR WPI; 2002-371829/40.  
XX PT Determining the degree of cytosine methylation in genomic DNA, useful for  
XX PT diagnosis and prognosis, comprises selective hybridization of amplicons  
XX PT from chemically treated DNA.  
XX PS Claim 12; 56pp + Sequence listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of  
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
XX CC genomic sample of DNA. The sample is treated chemically to convert  
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
XX CC DNA that contains the target C is amplified to form a labeled amplicon.  
XX CC The amplicon is hybridized to two classes, each with at least one member,  
XX CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
XX CC degree of hybridization to both classes is determined from the label on  
XX CC the amplicon. From the ratio of labels hybridized to the two classes of  
XX CC oligomers, the degree of methylation is calculated. The method is used:  
XX CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
XX CC and of a wide range of diseases, e.g. cancer, disorders of the central  
XX CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
XX CC particularly by detecting mutations or single nucleotide polymorphisms  
XX CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
XX CC investigating cell differentiation. The method allows the methylation  
XX CC status of many C residues to be determined simultaneously. ABQ13410-  
XX CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
XX CC for determining the degree of cytosine methylation described in the  
XX CC disclosure of the invention  
XX  
SQ Sequence 631 BP; 85 A; 70 C; 256 G; 220 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 3.68e-05 Length: 631  
Score: 120.00 Matches: 21  
Percent Similarity: 75.00% Conservative: 0  
Best Local Similarity: 75.00% Mismatches: 7  
Query Match: 77.92% Indels: 0  
DB: Gaps: 0  
  
US-10-010-050A-2_COPY_1_28 (1-28) x ABQ13702 (1-631)  
QY 1 MetAArgGlyAlaAlaArgGlyArgAlaSerTPCyGTTTAlaLeuAlaLeu 20  
DB 209 ATGCGCGGGGCGCGGGCGGCTTTTGGTGTGGGTTTGGGCTTG 268  
|||  
QY 21 LeuTTPLeuAlaValValProGly 28  
|||
```


CC nucleic acids involved in bronchoconstriction, allergies, and inflammation. The ON can have anti-inflammatory effects.

06-APR-1999; 99US-0127958P.

XX

Sun Apr 25 18:11:42 2004

us-10-010-050a-2_copy_1_28.rng

Page 10

Search completed: April 24, 2004, 23:26:32
Job time : 55.6833 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2004, 23:02:32 ; Search time 4391.8 seconds

(without alignments)
2352.639 Million cell updates/sec

Title: US-10-010-050A-2
Perfect score: 1927
Sequence: 1 MRGAGARARASMCWALAL.....IKTYERIPRINKITSL 346

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

searched: 2751329 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+g2n.model -DEV=xlh
Q=/cgm_1/USPTO.spool/US10010050/runat_22042004_113204_27556/app_query.faste_1.1372
DB=EST-QPMT=fastcap-SUFFIX=txt-MINMATCH=0.1-LOOPCH=0-7LOPEXT=0
UNITS=bits-START=1-END=1-MATRIX=blomum62-TRANS=human40.cdi-LIST=45
DOALIGN=200-THR SCORE=pct-THR MAX=100-THR MIN=0-ALIGN=15-MODE=LOCAL
OUTFMT=ptc-NORM=ext-HEAPSIZE=500-MINLEN=0-MAXLEN=2000000000
USER=US10010050.@CCN 1.1.6283.@runat_22042004_113204_27556-NCPU=6-ICPU=3
NO_MMAP-LARGESUBSTRY-NB6-SCORES=0-WAIT-DSPBLOCK=100-LONGLOG
DEV TIMEOUT=120-WARN TIMEOUT=30-THREADS=1-XGAPOP=10-XGAPEXT=0.5-FGAPOP=6
FGAPEXT=7-YGAPOP=10-YGAPEXT=0.5-DELOP=6-DELEXT=7

Database: EST:
1: em_estba:*
2: em_estbm:*
3: em_estin:*
4: em_esttm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_esttm:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_in:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pig:*
27: em_gss_vit:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1739	90.2	1201	9	AL546472
2	1451.5	75.3	2356	11	AK043247
3	1443.5	74.9	2433	11	AK085741
4	1429	74.2	2247	11	AK032293
5	1394	72.3	923	13	BX331615
6	1386.5	72.0	924	14	CA488543
7	1360	70.6	869	12	BI090566
8	1312	68.1	787	12	BI223533
9	1280.5	66.5	1077	13	BX442662
10	1245	64.6	1058	10	BF982158
11	1237	64.2	689	13	BU621797
12	1214	63.0	1201	9	AL571805
13	1211	62.8	931	10	BE873363
14	1207	62.6	641	14	CB215456
15	1187.5	61.6	974	13	BQ319385
16	1167	60.6	908	13	BX370641
17	1158	59.7	915	12	BI080303
18	1149.5	59.1	938	13	BQ926170
19	1128	58.5	676	12	BG820113
20	1113	57.8	755	13	BQ443453
21	1099.5	57.1	777	10	AM475730
22	1097.5	57.0	872	12	BI734549
23	1073.5	55.7	828	12	BI156421
24	1070	55.5	607	12	BM849262
25	1056	54.8	833	14	CK182221
26	1054.5	54.7	953	12	BI456284
27	1018.5	52.9	915	13	BX757396
28	1012	52.5	564	14	CB297292
29	1012	52.5	623	13	BQ417448
30	996.5	51.7	796	12	BI182240
31	989	51.3	555	9	AU280382
32	981.5	50.9	650	14	CF724176
33	969.5	50.3	867	13	BX757383
34	952	49.4	868	14	CF593022
35	951	49.4	767	9	AJ451053
36	948.5	49.2	854	12	BI082103
37	922	47.8	869	14	CF378929
38	910.5	47.2	661	13	BY734042
39	900	46.7	922	14	CA789305
40	899	46.7	607	28	AQ309525
41	898	46.6	847	13	BQ719074
42	893	46.3	720	12	BI111605
43	886	46.0	736	12	BI077452
44	881.5	45.7	706	10	BF739857
45	880	45.7	516	14	CB455600

ALIGNMENTS

RESULT 1
LOCUS AL546472 1201 bp mRNA linear EST 31-MAY-2003
DEFINITION AL546472 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
Accession AL546472
VERSION AL546472.2 GI:31268306
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1201)

DEFINITION Mus musculus 10 days lactation, adult female mammary gland cDNA, RIKEN full-length enriched library, clone: D730033P03 product: similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5 PROTEIN) [Homo sapiens], full insert sequence.

ACCESSION AK085741.1 GI:26102938

VERSION AK085741.1

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)

JOURNAL MEDLINE 99279253

PUBMED 10349636

REFERENCE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL MEDLINE 20499374

PUBMED 11042159

REFERENCE Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimura, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, S., Kawai, J., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, I., Matsubara, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL MEDLINE 20530913

PUBMED 11076861

REFERENCE The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)

JOURNAL MEDLINE 11076861

PUBMED 11076861

REFERENCE The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)

JOURNAL MEDLINE 11076861

PUBMED 11076861

REFERENCE Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirokane, T., Hori, F., Imoto, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, C., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

FEATURES

SOURCE

1..2433

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:D730033P03"

/db_xref="MGI:242937"

/db_xref="taxon:10090"

/clone="D730033P03"

/sex="female"

/tissue_type="mammary gland"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="10 days lactation, adult"

1..2433

/note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5 PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence: PASTY, 73.9%ID, 86.7%length, match=1041)"

ORIGIN

misc_feature

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:
2,04e-134	1443.50	83.62%	74.71%	11
2433	260	31	48	3

US-10-010-050a-2 (1-346) x AK085741 (1-2433)

1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrp-----CysTrpAlaLeu 18

132 ATGCTGCGGGGGGGGGCAGTGGG-----GGCAGTGGGGGGGGGGGGCTGGCTG 182

19 AlaLeuLeuTrpLeuAlaValAlaProGlyTTPSerArgValSerGlyLeuProSerArg 38

183 GGGCTGCTGGGGCTGGCGAGCATTCGGGGCGGCTCCCGACGCTGGG----- 230

39 ArgHisTrpProValProTyrTyrArgPheArgPheArgPheArgProArgProArgProArg 58

231 CAAGCTGAGCGGGTGGCGTCAAGCGCTTCTCTTCCGTCGGAAGACAGATCCCTCTGT 230

59 GlnAlaLeuTyrTrpPheCysProThrGlySerProIleProValMetGlnGlyAspAsp 78

291 CAAGCTAAGTATATCTTCTGCTTACCGGCTGCCATCCGATTATGAAGACATGAC 350

79 AspIleGluValPheArgLeuGlnAlaProValTTPGluPheLeuTyrGlyAspLeuLeu 98

351 GTCATGAGAGTCTTACGACTACCAAGCCCGATTGGGAAATTAAATGAGACCTCTCG 410

99 GlyHisLeuLeuTyrLeuMetHisAspAlaIleGlyPheArgSerThrLeuTrpGlyLysAsn 118

411 GGCACATTTAAATTAATGATGATGAGCGCGGGGATTGAGAGCACACTGACAGGCAAGAC 470

119 TyrThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArg 138

471 TACACATTTAGAGTATGATGACTTTCCAGCTGGGCACTGATCATTTCCCACTCCGG 530

139 ProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGlnGlyLeu 158

531 CCGACAGAGGCGCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 550

159 AspAspValHisTrpTyrGluLeuGlnGlyTrpLeuValGlnValAlaThrIleSerGlyAsn 178

591 GATGATTAACACTGGAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 650

179 MetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGlnTrpGlyTyrTyrGlu 198

651 ACATTTAACAAAGTGGCGAGTGGGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 710

Percent Similarity: 86.50% Conservative: 31
 Best Local Similarity: 76.99% Mismatches: 40
 Query Match: 74.16% Indels: 4
 3: 11 Gaps: 1

3-10-010-050a-2 (1-346) x AK032293 (1-2247)

```

Y 21 LeuTrpLeuAlaValAlaProGlyTyrSerArgValSerGlyIleProSerArgArgHis 40
C CTCGGCGTGGGAGACATTCCTGGGCGGCGGCCGCCAGCTGGG-----CAACGC 50
Y 41 TrpProValProTyrLysArgPheAspPheArgProLysProAspProTyrCysGlnAla 60
C TGGCGGCGGCGCATACAGACACTTCTCTTCCGCGGAAGACAGATCCCTACTGCAAGCT 110
Y 61 LysTyrTrpPheCysProThrGlySerProIleProValMetGluGlyAspAspAspIle 80
C LysTyrTrpPheCysProThrGlySerProIleProValMetGluGlyAspAspAspIle 170
Y 81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAspLeuLeuGlyHis 100
C GAGGCTTACAGACTACAGACCCCGATTTGGGAATTTAAATTTGAGACCTCTGGGACAC 230
Y 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
C TTAAACTTATGACATGACGCGCGGATTCAGAGACACATGACAGGCAAGAACCTACACA 290
Y 121 MetGluTyrPyrGluLeuPheGlnLeuGlyAsnGlySerThrPheProHisLeuArgProGlu 140
C ATTGAGTGTATGAACTTTCCAGCTGGGCAACGTACATTTCCACCTCGGCGCTGAC 350
Y 141 MetAspAlaProPheTrpCysAsnGlnGlyAlaIaCysPhePheGluGlyIleAspAsp 160
C AAGAGCGCTCCCTTCTGATGTAACCAAGGGGAGCGCTCTTTTGAAGAAATGATGAT 410
Y 161 ValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180
C AAACACTGGAAGAAACGCGGACACTGTCAGTCGTCGACACCATATCCGGAACACATTT 470
Y 181 AsnGlnMetAlaLysTrpValLysGlnAspAsnGluTrpGlyIleTyrTyrGluThrTrp 200
C AAAAAAGTGGCGGAGTGGTGAAGAGACATGAACTGGGATTTATTATGACATG 530
Y 201 AsnValLysAlaSerProGluLysGlyAlaGluTrpTrpPheAspSerTyrAspCysSer 220
C AAGGTCGAGCGCGGCGGACGACAGACAGGGGCCAGACTGCTTCAAGTCTTCAAGCTGTCG 590
Y 221 LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240
C AATTTTGTCTTAAAGACATATGAAATTTGGCTGAAATTTGGAAACGAATTCAGAGAGATA 650
Y 241 GluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGlu 260
C GAAACCAACTATACAAATATTTCTTACAGTGAAGGCCATTTACTCTGGGAATGAA 710
Y 261 ThrSerValPheGlyProThrGlyAsnLysPheLysGlyLeuAlaIleLysArgPheTyr 280
C ACATCATATTTTGGGCGGCAAGAAACAGACTCTTGGCTTGGCCATTAATAAATTTAT 770
Y 281 TyrProPheLysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePhe 300
C GGGCCCTTCAAGCGATTTGTCAACCAAAATTTCTGATGATAAATTTCTTGAATAATTTT 830
Y 301 AspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu 320
C GATACGATATTTATACACAGACAGCTTCTTCTTTTATACCTTGAAGATGGTTTCA 890
Y 321 PrometLysPheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsn 340
C CCAATGAACCCCTTGTCAAAATATACATGAGAAACCCGTTTACCTACCGGACAT 950
Y 341 LysThrLeuSerGlyLeu 346
  
```

Db 951 ACAACATTTACGACTTG 968

RESULT 5
 BX331615
 LOCUS
 DEFINITION
 CDNA clone CS0DB009YCO1 5-PRIME, mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 923)
 L.W.B., Gruber,C., Jesse,J., and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 COMMENT
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 7238.f for
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DB009AB01QPI&cluster=7238.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/InvitrogenCorporation1600
 Parady Avenue Genoscope sequence ID : CS0DB009AB01QPI.
 Location/Qualifiers
 1..923
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DB009YCO1"
 /issue_type="NEUROBLASTOMA COT 10-NORMALIZED"
 /clone_id="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-cligo (dt)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
 Pred. No.: 5.55e-130 Length: 923
 Score: 1394.00 Matches: 251
 Percent Similarity: 94.36% Conservative: 15
 Best Local Similarity: 94.36% Mismatches: 15
 Query Match: 72.34% Indels: 0
 DB: 13 Gaps: 0

US-10-010-050a-2 (1-346) x BX331615 (1-923)

```

QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTyrCysTrpAlaLeuAlaLeu 20
Db 118 ATGGCGCGGCGGCGGCGGCGGCTCGGGAGCGGCTTCTGATGCTGGCCCTGGCGCTG 177
QY 21 LeuTrpLeuAlaValAlaProGlyTyrSerArgValSerGlyIleProSerArgArgHis 40
Db 178 CTGGGCGGCGGCGGCTTCCGGGCTGTCGCGGCTCGGCGATCCCTCCGGGCGCAC 237
QY 41 TrpProValProTyrLysArgPheAspPheArgProLysProAspProTyrCysGlnAla 60
Db 238 TGGCGGCGGCGGCGGCGGCGGCTTCAAGCTTCCGTCACAAACCTGATCTTATTTGCAACT 297
QY 61 LysTyrTrpPheCysProThrGlySerProIleProValMetGluGlyAspAspAspIle 80
Db 298 AAGTATACCTTCTGTCCAACTGCGCTCACCTTCCAGTTATGAGAGCGGTGATACATT 357
QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAspLeuLeuGlyHis 100
Db 358 GAACTTTTCAATTACAGACCCCGATATGGGAATTTAAATATGAGACCTCTGGGACAC 417
  
```

101 LeuLysellIewethIsaPaalalleGlyPheAArgSerThreUthrGlyIySaenYrThr 120
 418 TTGAAAATTATGATGNTGNCATTGATTCAGAGTACATTAACGGAGAACTANACA 477
 121 MetGluTrpYrGluLeuPheGlnLeuGlyAsnCySerThreProHisLeuArgProGlu 140
 478 ATGGATGTATTAACCTTTCCAACTTGGGAANTGTANATTCCCNCTCCGNNCTGNA 537
 141 MetAspAlaProPheTrpCysAengInGlyAlaAlaCysPhePheGluGlyIleAspAsp 160
 538 ATGGRTGSSCCTTTCTGKGATATCAAGGGGTGCTGCTTTTGGGGAAATGRTGAT 597
 161 ValHisTrpYrGluAengInGlyTrpLeuValGlnValAlaThrIleSerGlyAanMetPhe 180
 598 GTTCACCTGGAAGAAATGGACATTAAGTCAAGTACCACTATACAGAAACATGTTTC 657
 181 AspGlnMetAlaIysTrpValIyGlnAspAsnGluThrGlyIleYrYrGluThrTrp 200
 658 AACCAATGGCAAGTGGTGGAAACAGACAAATGAACAGAAATTTATTAAGACATGG 717
 201 AsnValIySaIaseProGluIyGlyAlaGluThrTrpPheAspSerYrAspCysSer 220
 718 AATGTAAACCCAGCCCAAGAAAGGGGCGAGACATGTTGATTCCTACGACTGTCC 777
 221 LysPheValLeuArgTrpPheAsnLysLeuAlaGlnPheGlyAlaGluPheLysAsnIle 240
 778 AATTTGTGTTAAGACCTTTAACAAGTTGCTGAATTTGAGACAGAGTTCAGAACATA 837
 241 GluThrAsnYrThrArgIlePheLeuYrSerGlyGluProThrYrLeuGlyAsnGlu 260
 838 GAACCAACTATACAGAAATATTTCTTTACAGTGGAGAACCTTACTTCGGAAATGAA 897
 261 ThrSerValPheGlyPro 266
 898 ACATCTTTTGGGCCA 915

SULT 6
 488543
 CUS
 FINITION
 CA488543 924 bp mRNA linear EST 14-NOV-2002
 AGENCOURT 10808884 MAPCL Homo sapiens cDNA clone IMAGE:6720309 5',
 mRNA sequence.
 CESSION
 KSION
 CA488543.1 GI:24950702
 EST.
 YWORDS
 DRCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 924)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 UNPUBLISHED (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: gsgabs-remail.nih.gov
 Tissue Procurement: Kristi A. Egland, Ira Pastan
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LHAM14279 row: j column: 21
 High quality sequence stop: 637.
 Location/Qualifiers
 1. .924
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6720309"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 hTERT-HMEL, LNCaP"
 /lab_host="EMD10B"

ORIGIN
 Alignment Scores:
 Pred. No.: 3,17e-129 Length: 924
 Score: 1386.50 Matches: 258
 Percent Similarity: 88.63% Conservative: 7
 Best Local Similarity: 86.29% Mismatches: 21
 Query Match: 71.95% Indels: 13
 DB: 14 Gaps: 3

US-10-010-050A-2 (1-346) x CA488543 (1-924)

1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeu 20
 16 ATGTGGCGGGCGCGGGCGGCTCGGGACGCGCTTCTGCTGCTGGGCGCTGGCGCTG 75
 21 LeuTrpLeuAlaValValProGlyTrpSerArgValSerGlyIleProSerArgArgHis 40
 76 CTTTGGCTGGCGGTGTTCCGGGCTGCTCGGGGTCTGGGCAATCCCTCCGGCGGCAC 135
 41 TrpProValProCylTrpLysArgPheAspPheArgProLysProAspProYrCysGlnAla 60
 136 TGCGCGGTGGCCCAAGAGGTTTGACTTCGTCGCAAAACCTGATCCTTATGCAAGCT 195
 61 LysTrpThrPheCysProThrGlySerProIleProValMetGluGlyAspAspIle 80
 196 AAGTATATCTTTCTGTCCAACTGGCTCACTATCCAGTTATGAGGAGTATGATGACATT 255
 81 GlnValPheArgLeuGlnAlaProValTrpGluPheLysTrpGlyAspLeuLeuGlyHis 100
 256 GAAGTTTTCGATTACAGCCCAAGTATGGAAATTTAATATGAGACCTCTGGGACAC 315
 101 LeuLysellIewethIsaPaalalleGlyPheAArgSerThreUthrGlyIySaenYrThr 120
 316 TTGAAAATTATGATGNTGNCATTGATTCAGAGTACATTAACGGAGAACTANACA 375
 121 MetGluTrpYrGluLeuPheGlnLeuGlyAsnCySerThreProHisLeuArgProGlu 140
 376 ATGGAATGTATTAACCTTTCCAACTTGGGAANTGTANATTCCCNCTCCGNNCTGNA 435
 141 MetAspAlaProPheTrpCysAengInGlyAlaAlaCysPhePheGluGlyIleAspAsp 160
 436 ATGGATGCCCCCTTTCTGKGATATCAAGGGGTGCTGCTTTTGGGGAAATGRTGAT 495
 161 ValHisTrpYrGluAengInGlyTrpLeuValGlnValAlaThrIleSerGlyAanMetPhe 180
 496 GTTCACCTGGAAGAAATGGACATTAAGTCAAGTACCACTATACAGAAACATGTTTC 555
 181 AsnGlnMetAlaIysTrpValIyGlnAspAsnGluThrGlyIleYrYrGluThrTrp 200
 556 AACCAATGGCAAGTGGTGGAAACAGACAAATGAACAGAAATTTATTAAGACATGG 615
 201 AsnValIySaIaseProGluIyGlyAlaGluThrTrpPheAspSerYrAspCysSer 220
 616 AATGTAAACCCAGCCCAAGAAAGGGGCGAGACATGTTGATTCCTACGACTGTCC 675
 221 LysPheValLeuArgTrpPheAsnLysLeuAlaGlnPheGlyAlaGluPheLysAsnIle 240
 676 AATTTGTGTTAAGACCTTTAACAAGTTGCTGAATTTGAGACAGAGTTCAGAACATA 735
 241 GluThrAsnYrThrArgIlePheLeuYrSerGlyGluPro---ThrYrLeuGlyIy 259
 736 GAACCAACTATACAGAAATATTTCTTTACAGTGGAGAAACCTTACTTATCCGGGGA 795

/clone lib="MAPCL"
 /note="Vector: PCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Egland, James J. Vincent, Robert Strausberg,
 Bunkook Lee & Ira Pastan: Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

Y 259 snGluThrSer-----ValPheGly-----ProThrGlyA 269
 b 796 AATTGAAACATTCGATTTTGGGGCTAACAGAACCAAGAACTCTTGAGGT 855
 Y 269 snLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHis 286
 b 856 CTAAACCCATTACAAAAAAATTTTAATTAACCCGCTTCCTCAAAACAC 908
 RESULT 7
 I090566
 ECUTION 602855673p1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996891 5',
 mRNA sequence.
 CCESION BI090566
 ERSION BI090566.1 GI:14508896
 EYMORDS EST
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNI at:
 http://image.llnl.gov
 Plate: L1AM1023 row: a column: 20
 High quality sequence stop: 843.
 Location/Qualifiers
 1..869
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4996891"
 /cell_line="MGC36"
 /lab_host="DH10B"
 /clone_1ib="NIH_MGC_10"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.5 kb. Library prepared by Life
 Technologies."
 RIGIN
 Alignment Scores:
 red. No.: 1,386-126 Length: 869
 core: 1360.00 Matches: 262
 Percent Similarity: 94.58% Conservative: 0
 est Local Similarity: 94.58% Mismatches: 8
 Query Match: 70.58% Indels: 8
 Gaps: 0
 US-10-010-050a-2 (1-346) x BI090566 (1-869)

Db 232 AAGTATACCTTCTGTCACCTGCTCACTATCCACTATGAGGGGTGATGACATT 291
 Qy 81 GluValPheArgLeuGlnAlaProvalTTrpGluPheLysTyrGlyAspLeuLeuGlyHis 100
 Db 292 GAGCTTTTGGATTACAGCCCAAGTATGGGAATTTAAATGAGAGCTCTCGGACAC 351
 Qy 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
 Db 352 TTGAAAATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411
 Qy 121 MetGluTyrPyrGluLeuPheGlnLeuGlyAsnCySerThrPheProHisLeuArgProGlu 140
 Db 412 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 471
 Qy 141 MetAspAlaProPheTyrCyAsnGlnGlyAlaAlaIleAspPhePheGlnGlyLysAsp 160
 Db 472 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 531
 Qy 161 ValHisTyrPyrGluLeuGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180
 Db 532 GTTCACTGGAAGAAATGGGACATGATGATGATGATGATGATGATGATGATGATGAT 591
 Qy 181 AsnGlnMetAlaLysTyrPyrValGlyGlnAspAsnGluThrGlyIleTyrTyrGluThrTyr 200
 Db 592 AACCAATGGC-AAGTGGGTAACAGACATGAAACAGAAATCTATATGACATGG 650
 Qy 201 AsnValLysAlaSerProGluLysGlyAla--GluThrTyrPheAspSerTyrAspCys 220
 Db 651 AATGTAAAGCCACCCAGAAACAGGGGCGACGAGACATGTTGATCTCAACGCTGT 710
 Qy 220 eLysPheValLeuArgThrPheAsnLysLeuAlaGlu-PheGlyAlaGluPheLysAsn 239
 Db 711 CCAATTTGTGTAAAGACCTTTAAACAAGTTGCTGAACCTGGAGACGAGTCAAGAG 770
 Qy 240 Ile-GluThrAsnTyrThrArgIlePheLeuTyrSerGly-GluProThrTyrLeuGly- 258
 Db 771 ATTGAAACCAANTATACAAATATTTCTTTAAGAGGAAACCTTATCTTCTGGAA 830
 Qy 259 AsnGluThr-SerValPheGlyProThrGlyAsnLys 270
 Db 831 AATGAAACCACTCTTTAAGGGGACCAACGAAACAG 867
 RESULT 8
 BI223533 787 bp mRNA linear EST 11-JUL-2001
 LOCUS 602941923p1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5105093 5',
 DEFINITION mRNA sequence.
 ACCESION BI223533
 VERSION BI223533.1 GI:14676977
 KEYWORDS EST
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNI at:
 http://image.llnl.gov
 Plate: L1AM1253 row: n column: 06
 High quality sequence stop: 762.
 Location/Qualifiers
 1..787
 /organism="Homo sapiens"
 source


```

355 TGAAGTTTTCGATTACAGCCCGAGTATGGGAATTAATATGAGACCTCTGGAGCA 414
100 sLeuylsllEmethisAspAlailegIlyPheAsySerThrLeuThrGlyysAsnYrTh 120
415 CTTGAAAAATATGATGATGAGCCAGTGAATTCAGAAATCAATTAATCTGCAAGAACTACAC 474
120 MetGluTrpIyrgIuLeuPheGlnLeuGlyAsnCyseThrPheProHisLeuArgProG 140
475 AATGGAATGATGATGAACTTTTCGAACTTGGCACTGACATCTCCCATCTCCGACTGA 534
140 uMetAspAlaProPheTrpCyseAsnGlnGlyAlaAlaCysePhePheGlnGlyIleAspAs 160
535 AAMGAGATGCCCTTCTGCTGCTGATTCARAGCGCTGCTGCTTTTAAAGGAAATGATGA 594
160 PValHisTrpIyrgIuLeuPheGlnLeuValGlnValAlaThrIleSerGlyAsnMetPh 180
595 TGTCACCTGAGAGAGAAATGGAGACATTAATGTCARAGCAATATCAAGAAACATGTT 654
180 eAsnGluMetAlaIystrPValIyGlnAspAsnGluThrGlyIleTyrgIuThrTr 200
655 CAACCAATGGCAAAAGTGGGAAACAGRCATGAAACAGRAATTTWTATTTAGACATK 714
200 PAsnValIyAlaIaserProGluIyGlyAlaGluThrTrpPheAspSerTyraSpCyse 220
715 GAATGCTMAAACARCCCAAAAAAGGGGAGAGACATGTTATTTATTTCTGACATGCTTC 774
220 rIyPheValleuArgThrPheAsnIySleuAlaGluPheGlyAlaGluPheIySAsnI 240
775 CAAMTTTGTATTAARGACCTTTAACAAATTTGCTGAATTTGACAGAGTTCAAPAAACAW 834
240 eGluThrAsnIyTrpThrArgIlePheLeuTySerGlyGluProThrTyrgIuLeuGlyAsnG 260
835 AAAAAACAATATACAAAATATTTCTTTTAAAGTGGARACT-ACTTATCTKGGGAATGA 893
260 uThrSerValPheGlyProThrGlyAsnIySleuThrIleuAlaIleTySArgPheTy 280
894 AACATCTGTTTGGGAGACARAGAAACMAAA---CTCTTGTCTTGCMTAAAAAAGATTW 950
280 rTyTrpPheIySProHisIleuProThrIyGluPheIleuSerLeuLeuGlnIlePh 300
951 TTASCCCTCMARCCCMWTTGSCAMHAAAAATKTTT-----TWAGKGYTTGAAMTTT 1004
300 eAspAlaValIleValHisIySglnPheTyLeu 311
1005 TTKBATTTATTTTGVMAAAMATKWTTTTTTT 1038

ESULT 10 1098 bp mRNA linear EST 23-JAN-2001
F982158 602308976P1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4400298 5',
EPINITIION mRNA sequence.
CCESION BF982158
EYWORDS BF982158.1 GI:12384970
EYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1098)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabds-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10105 row: 9 column: 19
High quality sequence stop: 693.

```

```

FEATURES
    source
        Location/Qualifiers
            1..1098
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4400298"
                /tissue="duodenal adenocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: small intestine; Vector: pCMV-SPORT6;
                Site:1: Notti; Site:2: Sali; Cloned unidirectionally;
                oligo-dT primed. Average insert size 1.767 kb. Library
                enriched for full-length clones and constructed by Life
                Technologies. Note: this is a NIH_MGC library."

ORIGIN
Alignment Scores:
Pred. No.: 7.61e-115 Length: 1098
Score: 1245.00 Matches: 240
Percent Similarity: 88.13% Conservative: 5
Best Local Similarity: 86.33% Mismatches: 30
Query Match: 64.61% Indels: 6
DB: Gaps: 2

US-10-010-050a-2 (1-346) x BF982158 (1-1098)
OY 1 MetArgArgGlyAlaGlyAlaAlaArgIyArgAlaSerTrpCystrPAlaLeuAlaIeu 20
Db 35 ATGGCGGGGGGGCGGGCGGGCGGCTCGGGAGCGGCTTCGTGCTGCGCTGGCGCTG 94
OY 21 LeuTrpLeuAlaValValProGlyTrpSerArgValSerGlyIleProSerArgHis 40
Db 95 CTTGGCTCGGGGGGTTCGGCGGGGTTCGGGCTCGGGGCTCGGGGCTCGGGGCTCGG 154
OY 41 TrpProValProTyrgIySArgPheAspPheArgProIySProAspProTyrgIySglnAla 60
Db 155 TGGCGGGGCTTCAAGACCGCTTGACTTCGTCGCAAACTGATCTTATGCAAGCT 214
OY 61 TyTrpThrPheCyseProThrGlySerProIleProValMetGluIyAspAspAspIle 80
Db 215 AAGTATCTTTCTGTCACATCGGCTCACTACCTATCCAGTTATGAGGGGATGATGACAT 274
OY 81 GluValPheArgLeuGlnAlaProValTrpGluPheIySglnIySglnIySglnIyS 100
Db 275 GAATGTTTTCATTAACAAGCCCGAGTGGAAATTAATTAATGAGACCTCTCGGACAC 334
OY 101 LeuylsllEmethisAspAlailegIlyPheArgSerThrLeuThrGlyysAsnYrThr 120
Db 335 TTGAAATATGATGATGATGAGCCCATTTGATTCAGAGTACATTAATCGCAAGAACTACACA 394
OY 121 MetGluTrpIyrgIuLeuPheGlnLeuGlyAsnCystrPheProHisIleuArgProGlu 140
Db 395 ATGGATGATGATGAACTTTCCAACTTGGCAACTGTACATTTCCCATCTCGACCTGAA 454
OY 141 MetAspAlaProPheTrpCyseAsnGlnGlyAlaAlaCysePhePheGlnGlyIleAspAsp 160
Db 455 ATGATGCCCCCTTCTGCTGCTGATTCAGAGCCGCTGCTGCTTTTAAAGGAAATGATAT 514
OY 161 ValHisTrpIyrgIuLeuPheGlnLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180
Db 515 GTTCACTGGAAGACACCTGGACATTAATCTTCAAGTACATTAATACGAAACATGTTCC 574
OY 181 AsnGluMetAlaIystrPValIyGlnAspAsnGluThrGlyIleTyrgIuThrTrp 200
Db 575 AACCAATGCAAAAGTGGTGAACAGACATGAAACAGAAATTAATCTGAGACCTGG 634
OY 201 AsnValIyAlaIaserProGluIyGlyAlaGluThrTrpPheAspSerTyraSpCyser 220
Db 635 CCGTCACAGCCAG-CCGAAACACGGGGCAGAGACCTGTTGATTCCTACGACGTTCC 693
OY 221 LysPheValIleArgThrPheAsnIySleuAlaGluPheGlyAlaGluPheIySAsnIle 240
Db 694 AATTT-GTGTACGGACCTT-AAACAAGTTGCTGATTTGGAGCAAGATCCAGAACCC-- 748

```


Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7238.f for more information about this cluster, see

http://www.genoscope.cns.fr/
cgl-bin/cluster.cgi?seq=CSODI030CE01NP1&cluster=7238.f. Contact :
Feng Liang Email : fliang@life.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Parade Avenue Genoscope sequence ID : CSODI030CE01NP1.

FEATURES

source

1.1201
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI030Y01"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ALIGN

Alignment Scores:

red. No. : 1.15e-111 Length: 1201
score: 1214.00 Matches: 230
Percent Similarity: 96.64% Conservative: 7
Best Local Similarity: 96.64% Mismatches: 3
Jery Match: 63.00% Gaps: 0

3-10-010-050A-2 (1-346) x AL571805 (1-1201)

110 PheargSerThleuThrGlyAsnTyrrThmetGluTrpTyrgluLeuphegInleu 129
1067 TTCAGAGTACATTACCTGCGAARAT-TACAAATTAATTTGGATACCA-ATTTCACACT 1010
130 GYAsnCythrPheProhileuArgProglumetAspAlaProPheTrpCysAsnGln 149
1009 GGCACATGATACATTTCCCATCTCCGACCTGAATGATSCCCTTCTGTATATCAA 950
150 GYAlAlAlAcysPhehegInleu-11ehapPheValhistrPlyGluAsnGlyThrie 163
949 GGGCGCGCCGCTTTTGTAGGAAATGATGATGTTCTGCTGAGAGAAATGGGCATT 893
169 uValGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTrpValIleG 183
889 AGTTCAAGTACGACATCATGAGAAACATGTTCAACCAATGGCAAGTGGGTGAAACA 833
189 naSPaAsnGluThrGlyIleTyrrTyrrThrThrPheAsnValIleSerProGluIle 203
829 GGCACATGAAACAGAAATTTATATGACATGAAATGTAAAGCCAGCCAGAAAGAG 773
209 yAlaGluThrTrpPheAsnSerTyrrAspCysSerIlePheValIleuArgThrPheAsn 229
769 GGCAGAGACATGATGATCTTACGACATGTTCAATTTGTGTAAAGACCTTAAACA 710
229 sleuAlaGluPheGlyAlaGluPheIleAsnIleGluThrAsnTyrrThrArgIlePhe 249
709 GTTGGCTGCAATTTGGACACAGTTCAAGACATAGAAACCACTATACAGAAATTTCT 680
249 uTyrrSerGlyGluProThrTyrrleuGlyAsnGluThrSerValPheGlyProThrGly 269
649 TTACAGTGAAGAACCTTACTATCTGGGAAATGAACAATCTGTTTGGGGCAACAGAAA 590
269 nlySThLeuGlyLeuAlaIleIleAsnArgPheTyrrTyrrProPheIleProhileu 289
589 CAAGACCTCTGTTTACCCATTAAGATTTATACCCCTTCAACCACTATTTGCCAAC 530
289 rlySGluPheIleuSerleuGlnIlePheAspAlaValIleValIleIleSGlnP 309
529 TAAAGATTTCTGTAGTGTCTCTGCAATTTTGTGATGAGATGATTTGTGCAAAACGTT 470
309 eTyrrLeuPheTyrrAsnPheGluTyrrPheLeuProMetIlePheProhileuIle 329

Db 469 CTATTTGTTTAAATTTTGAATTTGCTTTTTCCTATGAAATTCCTTTATTAAT 410
Cy 329 eThTyrrGluGluIleProleuProIleArgAsnIleThleuSerGlyLeu 346
Db 409 AACATATGAAGAAATCCCTTACTATCAGAAACAAACACTCTCTGTGTTA 358

RESULT 13
LOCUS BE873363 931 bp mRNA linear EST 20-OCT-2000
DEFINITION 601450425P1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3854119 5',
mRNA sequence.
ACCESSION BE873363
VERSION BE873363.1 GI:10322139
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@b-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://imgc.llnl.gov
Plate: L14M9579 row: b column: 08
High quality sequence stop: 662.

FEATURES

source

1.931
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3854119"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_65"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

ORIGIN

Alignment Scores:
red. No. : 1.65e-111 Length: 931
score: 1211.00 Matches: 232
Percent Similarity: 93.57% Conservative: 14
Best Local Similarity: 93.17% Mismatches: 4
Query Match: 62.84% Gaps: 1
DB: 10

US-10-010-050A-2 (1-346) x BE873363 (1-931)

77 AsPaSPaSP1leGluValPheArgLeuGlnAlaProValTrpGluPheIleTyrrGlyAsp 96
2 GATGATGACATTTGAAGTTTTCATTAAGAGCCAGTATGGGAAATTAAATATGAGAC 61
97 leuLeuGlyValIleuValIleMetChIleAspAlaIleGlyPheArgSerThrleuThrGly 116
62 CTCCTGGACACTTGAATAATATGATGATGATGATGATGATGATGATGATGATGATGAT 121
117 lyeAsnTyrrThmetGluTrpTyrrGluLeuPheGlnLeuGlyAsnCythrPheProhile 136
122 AAGAACTACACATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181
137 leuArgProGluMetAspAlaProPheTrpCysAsnGlnIleValAlaCysPhePheGlu 156

WORDS EST.
ORCE Homo sapiens (human)
ORGANISM Homo sapiens
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 974)
 NIH-MGC <http://mgs.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNML)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNML at:
<http://image.llnl.gov>
 Plate: ILNML2574 row: 1 column: 13
 High quality sequence stop: 634.

FEATURES **source**

1. 974
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6382620"
 /tissue_type="cardioma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_11b="NIH_MGC_40"
 /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGCG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

Alignment Scores:
 ted. No.: 4.12e-109 Length: 974
 core: 1187.50 Matches: 228
 Percent Similarity: 78.77% Conservative: 28
 Percent Local Similarity: 70.15% Mismatches: 47
 Jery Match: 61.62% Indels: 22
 Gaps: 6

S-10-010-050a-2 (1-346) x BQ919385 (1-974)

Y 4 GYALAGLYALALAGGLY-----ArgAlaserTrp-----CystrpAlaleu 18
 |||:::|||||
 b 15 GGATCGCGATGCTGCGCGCGCGCATGTGAGCGCACTGGCGCGCGCTCGCGCTG 74
 Y 19 AAlaleuLeuTrpLeuAlaValAlProGlyTrpSerArgValSerGlyLeuProSerArg 38
 |||:::|||||
 b 75 GCGCTCTCGGCTGCGCGCATGCTGCGCGCTGCCGCGCGCTGCG----- 122
 Y 39 ArgHisTrpProValProTrpLysArgPheAspPheArgProLysProAspProTrpCys 56
 |||:::|||||
 b 123 CAACGCTGCGCGCTGCGCGCATGCGCTCTCTTCCGTCGGAAGACAGATCCCTACTGT 182
 Y 59 GlnAlaLysTrpThrPheCysProThrGlySerProLysPheProValMetGluLysAsp 78
 |||:::|||||
 b 183 CAAGCTAAGTAACTTCTGCTTACCGGCTGCCATCCAGTTAAGAGCAATGAC 242
 Y 79 AspIleGluValPheArgLeuGlnAlaProValTrpGluPheLysTrpGlyAspLeu 98
 |||:::|||||
 b 243 GTCATGAGGTCTTACGACTACAGCCCGATTTGGAAATTAAATATGAGACCTCTCG 302
 Y 99 GLYHisLeuLysTrpLeuMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsn 113
 |||:::|||||
 b 303 GGACACTTAACTTAATGATGATGCGCGCTGCGATTCAGAGCACACTGACGCGCAAGAAC 362

QY 119 TyrThrMetGluTrpYrGluLeuPheGlnLeuGlyAsnCystrpPheProHisLeuArg 138
 |||:::|||||
 Db 363 TACCAATGAGTGTGATGAACTTTTCAGCTGGCAACTGTACATTTCCCACTCCG 422
 QY 139 ProGluMetAspIleProPheTrpCysAsnGlnGlyAlaIleAspPheGluGlyLe 158
 |||:::|||||
 Db 423 CCGTCAAGAGCGCTCCCTTCTGTGTGACCAAGGCGCGCTCTTTTGAAGGAATA 482
 QY 159 AspAspValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsn 178
 |||:::|||||
 Db 483 GATGATTAACCTCGAAGGAAAGGAGCACTGTACGTGTCACCATATCCGGAAC 542
 QY 179 MetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGluTrpGlyIleTrpGlu 198
 |||:::|||||
 Db 543 ACATTTAAACAAAGCGCGGAGTGTGAGCAAGCAATGAACTGGGATTTATTATGAG 602
 QY 199 ThrTrpAsnValLysAlaSerProGluLysGlyAlaGluTrpPheAspSerTrpAsp 218
 |||:::|||||
 Db 603 ACATGACGCTCCGAGCGCGCGCCAGACAGAGGCGCGAGCTGCTTGAATCTTACAC 662
 QY 219 CysSerLysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLys 238
 |||:::|||||
 Db 663 TGTTCGAATTTGTCTTAAGGACATATAGAAATTTGCTGATTTGGAAAGAAATTCAG 722
 QY 239 AsnIleGluTrpAsnTrpThrArgIlePheLeuTrpSerGlyGluProThrTrpLeuGly 258
 |||:::|||||
 Db 723 AAGATAGAAACAACTATACGAAATTTCTTACGTGAGAGCTTATTTACTGGA 782
 QY 259 AsnGluTrpSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle-LysA 278
 |||:::|||||
 Db 783 AATGAACATTTATTTTGGCCCAAGGAAACAAAGCTTGTGCTTGGCCATAAAAA 842
 QY 278 TrpPheTrpTrpProPheLysProHisLeu---ProThrLysGlu-PheLeuLeuSerLeu 296
 |||:::|||||
 Db 843 AATTTTATTTGGCCCTTCCAGACCGAATTTGTTCACCAAGATTTCTCTGATGAATTT 902
 QY 297 LeuGln---IlePheAspAlaVal-----IleValHisLys 307
 |||:::|||||
 Db 903 CTGAAAAATTTTGTATACGATGATTAACCCCGAGGCTTCTTAATCTGTATTAAA 962
 QY 308 GlnPheTrp 310
 |||:::|||||
 Db 963 ACCTTTGG 971

Search completed: April 25, 2004, 05:45:56
 Job time : 4408.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame_plus_pzn model

on: April 24, 2004, 23:04:07 ; Search time 689.597 Seconds
(without alignments)
2262.168 Million cell updates/sec

file: US-10-010-050A-2
effect score: 1.927
quence: 1 MRGAGAGAGAGACWALAL.....IKITVEFIPIPINKTILSL 346

ring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

arched: 2907579 seqs, 2254313464 residues

nal number of hits satisfying chosen parameters: 5815158

imum DB seq length: 0
imum DB seq length: 2000000000

at-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

mand line parameters:
DEL=frame+ pzn.model -DEV=xlh
=/cg2_1/USPTO.spool/US10010050/rnat 22042004.113205.27584/app query.fasta_1.1372
B=Published Applications NA -OFMT=fastap -SUFFIX=rmh -MINMATCH=0.1
DOPE=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
RANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=PCT -THR MAX=100
HR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
KALEN=2000000000 -USER=US10010050 @CGN 1.1 776 @rnat 22042004.113205.27584
CPU=6 -ICPU=3 -NO MAP -LARGOQUERY -NEG SCORES=0 -NAIT -DSBLOCK=100
JNGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5
XAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase: Published Applications NA:*

1: /cg2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cg2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cg2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cg2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cg2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cg2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cg2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cg2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cg2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cg2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cg2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cg2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cg2_6/prodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cg2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cg2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cg2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cg2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cg2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cg2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Query	ID	Description
Match Length DB		

1	1927	100.0	1486	9	US-09-122-383-1	Sequence 1, Appli
2	1927	100.0	1486	14	US-10-010-050A-1	Sequence 1, Appli
3	1916	99.4	1751	13	US-10-653-595-26	Sequence 26, Appl
4	1916	99.4	1751	13	US-09-397-945-26	Sequence 26, Appl
5	1631	84.6	1038	9	US-09-122-383-13	Sequence 13, Appl
6	1631	84.6	1038	14	US-10-010-050A-13	Sequence 13, Appl
7	886	46.0	494	9	US-09-864-761-22592	Sequence 22592, A
8	727	37.7	506	14	US-10-040-739-825	Sequence 825, App
9	684.5	35.5	474	9	US-09-864-761-5832	Sequence 5832, Ap
10	472	24.5	497	16	US-10-264-049-1859	Sequence 1859, Ap
11	359.5	18.7	351	9	US-09-867-701-2579	Sequence 2579, Ap
12	241	12.5	473	9	US-09-954-456-1363	Sequence 1363, Ap
13	121	6.3	60	10	US-09-908-975-10339	Sequence 10339, A
14	112	5.8	267	15	US-10-198-846-9753	Sequence 9753, Ap
15	108.5	5.6	3530	16	US-10-369-493-33507	Sequence 33507, A
16	100.5	5.2	1827	13	US-10-425-114-21585	Sequence 21585, A
17	97.5	5.1	750	16	US-10-369-493-32308	Sequence 32308, A
18	97.5	5.1	986	16	US-10-369-493-36401	Sequence 36401, A
19	96	5.0	870	16	US-10-369-493-26859	Sequence 26859, A
20	96	5.0	3035	16	US-10-108-260A-1791	Sequence 1791, Ap
21	96	5.0	3580	10	US-09-822-846-21	Sequence 159, App
22	96	5.0	3580	10	US-09-374-046A-159	Sequence 159, App
23	96	5.0	3580	13	US-10-616-263-159	Sequence 127, App
24	94.5	4.9	14070	9	US-09-870-759-127	Sequence 127, App
25	94.5	4.9	14070	10	US-09-751-708A-127	Sequence 1826, Ap
26	94.5	4.9	27189	13	US-10-087-192-1828	Sequence 22365, A
27	94	4.9	4170	13	US-10-262-839-91	Sequence 91, Appl
28	94	4.9	4170	13	US-10-262-839-91	Sequence 3297, Ap
29	94	4.9	30013	9	US-09-764-877-3297	Sequence 3297, Ap
30	94	4.9	30013	16	US-10-242-515-3297	Sequence 1, Appli
31	93	4.8	1826	15	US-10-028-245-1	Sequence 22, Appl
32	93	4.8	4683	9	US-09-764-868-1494	Sequence 1494, Ap
33	93	4.8	4683	9	US-10-114-170-22	Sequence 11, Appl
34	93	4.8	39178	17	US-10-210-589-11	Sequence 9, Appli
35	92.5	4.8	1612	14	US-10-003-392-9	Sequence 19, Appl
36	92.5	4.8	2115	14	US-10-003-392-9	Sequence 120013, A
37	92	4.8	1468	13	US-10-424-599-120013	Sequence 23914, A
38	92	4.8	1815	13	US-10-282-122A-23914	Sequence 49, Appl
39	92	4.8	3719	13	US-10-147-493-49	Sequence 49, Appl
40	92	4.8	3719	13	US-10-145-127-49	Sequence 49, Appl
41	92	4.8	3719	13	US-10-160-503-49	Sequence 49, Appl
42	92	4.8	3719	13	US-10-143-118-49	Sequence 49, Appl
43	92	4.8	3719	13	US-10-144-993-49	Sequence 49, Appl
44	92	4.8	3719	13	US-10-158-787-49	Sequence 49, Appl
45	92	4.8	3719	13	US-10-140-024-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1
US-09-122-383-1
Sequence 1, Application US/09122383A
Patent No. US2002042093A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38
CURRENT APPLICATION NUMBER: US/09/122,383A
CURRENT FILING DATE: 1998-07-24
EARLIER APPLICATION NUMBER: 60/053, 613
EARLIER FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1486
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (47)...(1084)

3-09-122-383-1

Alignment Scores:

Seq. No.:	1,18e-231	Length:	1486
Score:	1927.00	Matches:	346
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
	9	Gaps:	0

3-10-010-050A-2 (1-346) x US-09-122-383-1 (1-1486)

```

1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
47 ATGGCGCGGGCGGGCGGGCGGGCTCGGGAGCGCCCTCTCGTGGCTGGCGCTGGCGCTG 106
21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyLeuProSerArgArgHis 40
107 CTTTGGCTCGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 166
41 TrpProValProGlyTrpArgPheAspPheArgProLysProAspProGlyCysGlnAla 60
167 TGGCGGGCGGGCGGGCGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 226
61 LysTrpTrpPheCysProThrGlySerProLysProLysProLysProLysProLysPro 80
227 AAGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 286
81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTrpGlyAspLeuLeuGlyHis 100
287 GAAAGTTTTCATTCAGAGCCCGAGTATGGAGATTTAATATGAGACCTCTGGGACAC 346
101 LeuLysIleMetHisAspAlaIleGlyPheArgSerTrpLeuThrGlyLysAsnTrpThr 120
347 TTGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406
121 MetGluTrpTrpGluLeuPheGlnLeuGlyAsnGlySerTrpPheProHisLeuArgProGlu 140
407 ATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 466
141 MetAspAlaProPheTrpCysArgGlnGlyAlaGlyTrpPheAspSerTrpLysAsp 160
467 ATGGATGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 526
161 ValHisTrpLysGluAsnGlyTrpLeuValGluValAlaThrIleSerGlyAsnMetPhe 180
527 GTTCTCTGAGAGAAATGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 586
181 AsnGlnMetAlaLysTrpValLysGlnAspAsnGluTrpGlyIleTrpTrpGluTrp 200
587 AACCAATATGCAAGTGGTGAACAGCAATGAAACAGAAATTTATATGACATG 646
201 AsnValLysAlaSerProGluLysGlyAlaGluTrpPheAspSerTrpLysAsp 220
647 AATGTAAAGCCAGCCCAAGAAAGGGGGGAGAGACATGTTGATTTCCACACATGTTCC 706
221 LysPheValLeuArgTrpPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsn 240
707 AATTTTGTGTAAGACCTTTTACAGATTGCTGATTTTGGACAGATGTCAGAAACAT 766
241 GluTrpAsnTrpTrpArgIlePheLeuTrpSerGlyGluProThrTrpLeuGlyAsn 260
767 GAAACCAATATGCAAGTATTTCTTACAGTGGAGAACCTTATCTGGGAAATGAA 826
261 ThrSerValPheGlyProThrGlyAsnLysTrpLeuGlyLeuAlaIleLysArgPhe 280
827 ACATCTGTTTGGCGCAACAGAAACAGAACTTGGTTTACCTTAAATAAATTTAT 886
281 TyrProPheLysProHisLeuProTrpLysGluPheLeuLeuSerLeuLeuGlnIlePhe 300
887 TACCCCTTCAACACATTTTGGCACTTAAGATTTCTGTGAGTCTCTTGGCAATTTT 946
301 AspAlaValIleValHisLysGlnPheTrpLeuPheTrpAsnGluTrpPheLeu 320

```

```

Db 947 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1006
Qy 321 PrometLysPheProPheIleLysIleTrpTrpGluGluIleProLeuProLysArg 340
Db 1007 CCTATATAATCCCTTTATATATAATGATGATGATGATGATGATGATGATGATGAT 1066
Qy 341 LysTrpLeuSerGlyLeu 346
Db 1067 AAAACACTCTCTGTTT 1084

```

RESULT 2

```

US-10-010-050A-1
; Sequence 1, Application US/10010050A
; Publication No. US2002017362A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; FILE REFERENCE: 97-38C1
; CURRENT APPLICATION NUMBER: US/10/010, 050A
; PRIOR FILING DATE: 2002-03-26
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/053, 613
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (47)...(1084)
US-10-010-050A-1

```

Alignment Scores:

Pred. No.:	1,18e-231	Length:	1486
Score:	1927.00	Matches:	346
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050A-2 (1-346) x US-10-010-050A-1 (1-1486)

```

Qy 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
Db 47 ATGGCGCGGGCGGGCGGGCGGGCTCGGGAGCGCCCTCTCGTGGCTGGCGCTGGCGCTG 106
Qy 21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyLeuProSerArgArgHis 40
Db 107 CTTTGGCTCGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 166
Qy 41 TrpProValProGlyTrpArgPheAspPheArgProLysProAspProGlyCysGlnAla 60
Db 167 TGGCGGGCGGGCGGGCGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 226
Qy 61 LysTrpTrpPheCysProThrGlySerProLysProLysProLysProLysProLysPro 80
Db 227 AAGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 286
Qy 81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTrpGlyAspLeuLeuGlyHis 100
Db 287 GAAAGTTTTCATTCAGAGCCCGAGTATGGAGATTTAATATGAGACCTCTGGGACAC 346
Qy 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerTrpLeuThrGlyLysAsnTrpThr 120
Db 347 TTGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406
Qy 121 MetGluTrpTrpGluLeuPheGlnLeuGlyAsnGlySerTrpPheProHisLeuArgProGlu 140

```

```

b 407 ATGGATGGTATGAACTTTCCAACTTGGCACTGACATTTCCCATCTCCGACCTGAA 466
b 407 ATGGATGGTATGAACTTTCCAACTTGGCACTGACATTTCCCATCTCCGACCTGAA 466
y 141 MetAPAPAPProPheTTPCySaSnGnGlyAlaAlaCyPhePheGluGlyIleAspAsp 160
y 467 ATGGATGGCCCTTTTCTGGTATATCAAGCGCTGCTGTTTTTTGAGGAAATGATGAT 526
b 161 VALHSTPLPGluAsnGlyThirLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180
b 527 GTTCACTGGAAAGAAATGGGACATTAGTTCAGTAGACAACTATATCAGSAAACAATGTT 586
y 181 ASnGlnMetAlaLysTrpValLysGlnAspAsnGlnThyGlyIleTyTrpGluThirP 200
b 587 AACCAATGGCAAGGAGGAGTGAACGAGCAATGAAACGAAATTAATATAGACATGG 646
y 201 AsnValLysAlaSerProGluLysGlyAlaGluThirPheAspSerTyAspCysSer 220
b 647 AATGTAAAGCCAGCCCAAGAAAGGGGCGAGACATGTTGATTCCTACACATGTTCC 706
y 221 LysPheValLysArgThirPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240
b 707 AAATTTGTGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGAGACAGATTCAAGAACATA 765
y 241 GluThirAsnTyThirArgIlePheLeuTySerGlyGluProThirTyTrpLysGlnGly 260
b 767 GAACCACTATACAAAGAAATTTCTTAAGAGGAAACCTACTTATCTGGAAATGAA 826
y 261 ThirSerValPheGlyProThirGlyAsnLysThirLeuAlaIleLysArgPheTy 280
b 827 ACATCTGTTTGGGCGCAACAGSAAACAAGACTCTGTTGAGCCATAAAGATTTAT 885
y 281 TyrProPheLysProIleLeuProThirLysGluPheLeuLeuSerLeuLeuGlnIlePhe 300
b 887 TACCCCTCAAAACCAATTTGCCAATAAGAAATTTCTGTTAGTCTCTGCAAAATTTT 945
y 301 AspAlaValIleValHisLysGlnPheTyTrpLeuPheTyAsnPheGluThirPheLeu 320
b 947 GATGCGATGATGTGCAACAACGTTCTATTTGTTTAAATTTGAATATGGGTTTAA 1006
y 321 PrometLysPheProPheIleLysIleThirTyTrpGluGlnIleProLeuProIleArgAsn 340
b 1007 CCTATGAATTCCTCTTTTATTAATAACATATGAAGAAATCCCTTATCTATCAGAAAC 1056
y 341 LysThirLeuSerGlyLeu 346
b 1067 AAAACACTCTCTGGTTTA 1084

```

```

PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1520)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-10-010-050A-2
Alignment Scores:
Pred. No.: 3,73e-230 Length: 1751
Score: 1916.00 Matches: 344
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 2
Query Match: 99.43% Indels: 0
Gaps: 0
DB:
US-10-010-050A-2 (1-346) x US-10-653-595-26 (1-1751)
QY 1 MetARARAGLVALAGLYALAlaArgGlyARGLASerTPCySerPAlaLeuAlaLeu 20
DB 49 ATGCGCGGGGCGCGGGCGGGCTCGGGAGCGCGCTTCTGTGCTGGCGCTG 108
QY 21 LeuTPLeuAlaValAlaProGlyTyTrpSerArgValSerGlyIleProSerArgArgHis 40
DB 109 CTTTGGCTCGCGGTGTCGGGCTGGTCCCGGGCTCGGCGCATCCCTCCGCGCCAC 168
QY 41 TrpProValProTyTrpLysArgPheAspPheArgProLysProAspProTyTrpCysGlnAla 60
DB 169 TGGCGGGGCGCTTACAGGCGCTTACCTTCCTGCAAACTGATCTTATGTCAAGCT 228
QY 61 LysTyTrpPheCyProThirGlySerProIleProValMetGlnGlyAspAspAspIle 80
DB 229 AAGTATCTTCTGTGCCAAGCTGCTACCTATCCAGTTATGAGGGGTATATGATGACATT 288
QY 81 GluValPheArgLeuGlnAlaProValIleGluPheLysTyTrpGlyAspLeuLeuGlyHis 100
DB 289 GAAGTTTTCGATTCACAGCCCAAGATGGAAATTAATAATAGAGACCTCCGAGACAC 348
QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThirLeuThyGlyLysAsnTyTrp 120
DB 349 TTGAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
QY 121 MetGluTrpTyGluLeuPheGlnLeuGlyAsnCyTrpPheProIleLeuArgProGlu 140

```

409 ATGAAATGATGATGACCTTTCCAACTTGCACTGACATTTCCCATCTCCGACTGAA 463
141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGlnGlyLeuAsp 163
469 ATGGAATGCCCCCTTTGCTGCTGATTAACAAGCGCTGCTGCTTTTGAAGGAATGATGAT 523
161 ValHisTrpIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 183
529 GTTCACTGGAAGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 583
181 AsnGlnMetAlaLeuTrpValIleGlnAspAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 203
589 AACCAATGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 643
201 AsnValIleValIleSerProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
649 AATGTAATGCAAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 708
221 LysPheValIleValIleTrpPheAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 240
709 AATTTGCTGTAAGACCTTTTCAAGTTGCTGTAAGTTGCTGTAAGTTGCTGTAAGTTGCT 768
241 GlnThrAsnTrpIlePheLeuTrpSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
769 GAAMCCAACTATACAGAAATTTCTTTACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 828
261 ThrSerValPheGlnProTrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280
829 ACATCTGCTTTTGGGCAAG 888
281 TyrProPheLeuProIlePheLeuTrpIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
889 TACCCCTTCAACACACATTTGCAAGATTTGCTGTAAGTGAAGTGAAGTGAAGTGAAGTGA 948
301 AspAlaValIleValIleValIleValIleValIleValIleValIleValIleValIle 320
949 GATGCACTGATTTGCAAGATTTGCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1008
321 ProMetIlePheProPheIleValIleTrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
1009 CCAATGAAATTTCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1068
341 LysThrLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 346
1069 AAAACACTCTCTGCTTTA 1086

RESULT 4
US-09-397-945-26
Sequence 26, Application US/09397945
Publication No. US20030065139A1
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc. et al.
TITLE OR INVENTION: 95 Human secreted proteins
FILE REFERENCE: P202/P1
CURRENT APPLICATION NUMBER: US/09/397,945
CURRENT FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578

PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,581
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,577
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,563
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,313
PRIOR FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1520)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-09-397-945-26

Alignment Scores:
Pred. No.: 3,73e-230 Length: 1751
Score: 1916.00 Matches: 344
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 2
Query Match: 99.43% Indels: 0
Gaps: 0

US-10-010-050A-2 (1-346) x US-09-397-945-26 (1-1751)

QY 1 MetArgArgGlyValAlaGlyAlaAlaArgGlyArgGlySerTrpCysTrpAlaLeuAlaLeu 20
DB 49 ATCGGCGGGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 108
QY 21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyIleProSerArgValHis 40
DB 109 CTTGGCTCGCGGCTGCTCGGCGGCTGCTCGGCGGCTGCTCGGCGGCTGCTCGGCGGCGGCGG 168
QY 41 TrpProValProTrpTrpLysArgPheAspPheArgProLysProAspProTrpCysGlnAla 60
DB 169 TGCGCGGCTGCTCAAGCGCTTGAAGCTTCTGCTCAAGCGCTTCTGCTCAAGCGCTTCTGCT 228
QY 61 LysTrpTrpPheCysProTrpGlySerProIleProValMetGlnGlyLysAspAspIle 80
DB 229 AAGTATCTTCTGCTCAAGCGCTTCAAGCGCTTCAAGCGCTTCAAGCGCTTCAAGCGCTTCAAG 288
QY 81 GluValPheArgLeuGlnAlaProValTrpGlnPheLysTrpGlyAspLeuLeuGlyHis 100
DB 289 GAAGTTTTCGATTAACAAGCCCGAGTATGGAATTTAAATATGAGAGAGCTTCTGCGGAGAC 348
QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerTrpLeuTrpGlyLysAsnTrpTrp 120
DB 349 TTGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
QY 121 MetGlnTrpTrpGlnLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGln 140
DB 409 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 468

141 MetAspAlaProPheTTPCySaengInglYAlaAlaCyPhePheGluGlylleAspAsp 160
 469 ATGGATGCCCTTCTGTGTATTCAGAGGGCGCTGCTTTTGGAGGAATGATGAT 528
 161 ValHisTrpLysGluAengLYThrLeuValGlnValAlaThrLleSerGlyAsnMetPhe 180
 529 GTTCACGTGGAAGAAAATGGGACATTAGTTCAGAGTACCACTATATCAGGAACATCTTC 588
 181 AsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlylleTyrrYrGluThrTrp 200
 589 AACCAATGCGCAAGTGGTGAAGAACAGCAATGAAGAACTTATATATGAGACATGG 648
 201 AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTrpAspCysSer 220
 649 AATGTAAACCCAGCCCAAGAAAGGGGGAGAGACATGTTGATTCCTAACGACTGTTCC 708
 221 LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnLle 240
 709 AATTTGTGTAAAGACCTTAAACAAGTTGGCTGAATTTGGACGACAGTTCCAGAACATA 768
 241 GluThrAsnTrpThrArgLlePheLeuTyrSerGlyGluProThrTrpLysGlyAsnGlu 260
 769 GAAACCAACTATACAGATATATTTCTTACAGTGAAGAACCTATCTGCGGAATGAA 828
 261 ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaLleLysArgPheTyr 280
 829 ACATCTGTTTTGGGCCAAGCAAGAAACAAAGCTTGGTTAGCAAAAGATTTAT 888
 281 TyrProPheLysProHisLysLeuProThrLysGluPheLeuSerLeuLeuGlnLlePhe 300
 889 TACCCCTTCAACCAACATTTGGCCACTTAAGAAATTTCTGTGAGTCTCTTGCAAAATTTT 948
 301 AspAlaValLleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu 320
 949 GATCAGCTGATGTGGCAAAACAGTTCTATTTGTTTATATTTGAAATTTGGTTTTTA 1003
 321 PrometLysPheProPheLleLysLleThrTyrGlnGluLlePheLeuProLleArgAsn 340
 1009 CCAATGAAATTCCTTTTATTAATTAACAATGAAGAAATTCCTTATCACTATCAGAAAC 1063
 341 LysThrLeuSerGlyLeu 346
 1069 AAAACACTCTCTGCTTGA 1086

SULT 5
 -09-122-383-13
 Sequence 13, Application US/09122383A
 Patent No. US20020042093A1
 GENERAL INFORMATION:
 APPLICANT: Sheppard, Paul O.
 APPLICANT: Gilbertson, Debra G.
 TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
 TITLE OF INVENTION: CHROMOSOME 13
 FILE REFERENCE: 97-38
 CURRENT APPLICATION NUMBER: US/09/122,383A
 CURRENT FILING DATE: 1998-07-24
 EARLIER APPLICATION NUMBER: 60/053,613
 EARLIER FILING DATE: 1997-07-24
 NUMBER OF SEQ ID NOS: 19
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 13
 LENGTH: 1038
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Degenerate nucleotide sequence encoding zsig46
 OTHER INFORMATION: polypeptide of SEQ ID NO:2
 FEATURE:
 NAME/KEY: variation
 LOCATION: (1)...(1038)
 OTHER INFORMATION: n is any nucleotide
 -09-122-383-13

Alignment Scores:
 Pred. No.: 1,18e-194 Length: 1038
 Score: 1631.00 Matches: 287
 Percent Similarity: 83.19% Conservative: 0
 Best Local Similarity: 83.19% Mismatches: 58
 Query Match: 84.64% Indels: 0
 DB: 9 Gaps: 0

US-10-010-050A-2 (1-346) x US-09-122-383-13 (1-1038)

QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
 Db 1 ATGGMGMGNGNGCNGCNGCNGCNGMNGCNGMNGCNGMNGTGCTGTGGCNGTNGCNGYT 60
 QY 21 LeuTrpLeuAlaValValProGlyTTPSerArgValSerGlyLleProSerArgAlaHis 40
 Db 61 YTNVTGTYTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNG 120
 QY 41 TrpProValProTyrLysArgPheAspPheArgProLysProAspProTyrCysGlnAla 60
 Db 121 TGCCNGTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNG 180
 QY 61 LysTrpThrPheCysProThrGlySerProLleProValMetGluGlyAspAspAspLle 80
 Db 181 AARTAAACNTTGTGCGAACNGCNGMNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNG 240
 QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTrpGlyLysAspLeuGluHis 100
 Db 241 GAGGTNTTNGNTNGCNGCNGCNGCNGTNGGAGTTTAAATGAGGAGGAGGAGGAGGAGG 300
 QY 101 LeuLysLleMetHisAspAlaLleGlyPheArgSerThrLeuThrGlyLysAsnTrpThr 120
 Db 301 YTNAAATTAATGCAArgArgCNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNGCNGTNG 360
 QY 121 MetGluTrpTrpGluLeuPheGluLeuGlyAsnCysThrPheProHisLysLeuArgProGlu 140
 Db 361 ATGAGATGGTGAATGATTTTTCATYTNNGGAAATGACNTTTCNCAYTNGMCCNGAR 420
 QY 141 MetAspAlaProPheTTPCySaengInglYAlaAlaCyPhePheGluGlylleAspAsp 160
 Db 421 ATGAGAGCNCNTTGTGTGTAACAGGAGCNGCNGTGTGTTTGTGAGGAGGAGGAGGAG 480
 QY 161 ValHisTrpLysGluAengLYThrLeuValGlnValAlaThrLleSerGlyAsnMetPhe 180
 Db 481 GTNCAITGGAGARARARAYGCAACNTYTNNGCAGTNGCNGCNGTNGCNGTNGCNGTNG 540
 QY 181 AsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlylleTyrrYrGluThrTrp 200
 Db 541 AAYCARATGCGNARATGGGTNARCARGAYAAVGARACNGGNAHTHTATYAGARACNTGG 600
 QY 201 AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTrpAspCysSer 220
 Db 601 AAYGTNAARCGMNSNCNGARARAGGNGCNGARACNGTGTGTGAYGWSNTYAGATGWSN 660
 QY 221 LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnLle 240
 Db 661 AARTTGTNTNMGNACNTTAAAYAAAYTNGCNGARTTGGCNGCNGARTTAAATAAATHT 720
 QY 241 GluThrAsnTrpThrArgLlePheLeuTyrSerGlyGluProThrTrpLysGlyAsnGlu 260
 Db 721 GARACNAATYAAACNMGNATHTTYYTNTATYWSNGCNGARCCNATYNTYTNNGGNAAGAR 780
 QY 261 ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaLleLysArgPheTyr 280
 Db 781 ACNWSNGTNTTGGCNGCNGCNGAATAAACAATYTNNGCNGTNGCNGTNGCNGTNTTAY 840
 QY 281 TyrProPheLysProHisLysLeuProThrLysGluPheLeuSerLeuLeuGlnLlePhe 300
 Db 841 TATCANTTAAACNCAYTNGCNGCNGAAGARTTNTYTNMWSNTYNTYTNCAATHTTY 900
 QY 301 AspAlaValLleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu 320


```

407 ATGAAAGTATGAACTTTTCCAACTGGCACTGTACATTCCCATCTCGACCTGAA 466
141 MetAspAlaProPheTrpCysAsnGlnGlyAlaIaIaCysPhePheGluGlyIleAspAsp 160
467 ATGGATGCCCCCTTTCTGCTGTATCAATCAAGCCGCTGCTCTTTTGGAGGAATGATCAT 526
161 ValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180
527 GTTCACTGGAGGAAATGGGACATTAGTTCAGTAGCACTATATCAGAAACATGTTCC 586
181 AsnGlnMetAlaIleTrpValIleGlnAspAsnGluTrpGlyIleTyrTrpGluThrTrp 200
587 AACCAAAATGGCAAAATGGGTGTAACAGACAAATGAAACAGAAATTTATTAATGACATGG 646
201 AsnValIleValSerProGluIleGlyAlaGluThrTrpPheAspSerTyrAspCysSer 220
647 AATGTAAAGCCAGCCAGAAAGGGGCGCAGACATGTTGATTCCTGACGCTGTTCC 706
221 LysPheValIleuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240
707 AAATTGGTTAAGGACCTTTTAACTGAGTGGCTGATTTGAGCAGAGCTTCAAGACATA 766
241 GluThrAsnTrpThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGlu 260
767 GAAACCAACTATACAGAAATATTTCTTTACAGTGAAGAACTTACTTATCTGGGAATGAA 826
261 ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
827 ACATCTGTTTGGGCCCAACAGAAACAAAGACTTGTGTTTGAATAAAGATTTTAT 886
281 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuLeuGlnIlePhe 300
887 TACCCCTTCAACCAATTTGCCAATTAAGAAATTCGTTAAGCTCTGGAATTTT 946
301 AspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu 320
947 GATGAGGATTTGTGCAACAAAGTTCTATTTGTTTATTAATTTGAATATGGTTTTTA 1006
321 PrometLysPheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsn 340
1007 CTTAAGAAATTCCTTTTATTAATTAATTAATGAAGAAATCCCTTACTTACGAATC 1066
341 LysThrLeuSerGlyLeu 346
1067 AAAACACTCTCTGGTTTA 1084

```

```

RESULT 3
5-10-653-595-26
Sequence 26, Application US/10653595
Publication No. US20040048304A1
GENERAL INFORMATION:
APPLICANT: Ruben et. al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027P1C1
CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03
PRIOR APPLICATION NUMBER: US 09/397945
PRIOR FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01

```

```

PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1520)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-10-653-595-26
Alignment Scores:
Pred. No.: 3,73e-230 Length: 1751
Score: 1916.00 Matches: 344
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 2
Query Match: 99.43% Indels: 0
DB: 13 Gaps: 0
US-10-010-050A-2 (1-346) x US-10-653-595-26 (1-1751)
QY 1 MetArgArgGlyAlaGlyAlaIaIaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
DB 49 ATGGCGCGGGGCGCGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 108
QY 21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyIleProSerArgArgHis 40
DB 109 CTTTGGCTCGCGGTGCTTCGGGCTGTCGCCGGTTCGGGCAATCCCTCCCGGCCAC 168
QY 41 TrpProValProTyrLysArgPheAspPheArgProLysProLysProTyrCysGlnAla 60
DB 169 TGGCCGCTCCCTCAACAGCGCTTGTACCTCCGTGCAAAACCGATCTTATGTCAAGCT 228
QY 61 LysTyrThrPheCysProThrGlySerProIleProValMetGlnGlyAspAspIle 80
DB 229 AAGTATACCTTTCTGTCAACTGGCTCACCTATCCGATTAGGAGGTATGATGACATT 288
QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAspLeuLeuGlyHis 100
DB 289 GAAGTTTTCGATTACAGCCCGAGTAGGAAATTAATATGAAACCTCTCGGACAC 348
QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
DB 349 TTGAATATTTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
QY 121 MetGluTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGlu 140

```

b 409 ATGAAATGTATGAACTTTTCCAACTTGCACTGTACATTTCCCATCTCCGACTGAA 468
y 141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGlnGlyLeuAsp 160
b 469 ATGATGATCCCTTTCTGCTGTATCAAGGCGCTGCTGCTTTTATGAGGAATGTATGAT 528
y 161 ValHisTrpIysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180
b 529 GTTCACTGGAAGAAATGGACATTAGTTCAAGTACGACATATATCAGGAAACAGTTTC 588
y 181 AsnGlnMetAlaIysTrpValIlyGlnAspAsnGlnThrGlyIleTrpTrpGluTrp 200
b 589 AACCAATGGCAAGTGGTGAACAGGACATGAAACAGAAATTTATTTATGACATGG 648
y 201 AsnValIlyAlaSerProGlnIlyGlyAlaGluThrTrpPheAspSerTrpAspCysSer 220
b 649 AATGTTAAACCCAGCCCAAGAAAGGGGGCAGACATGTTTATTCACGACTGCTTC 708
y 221 LysPheValIleuArgTrpPheAsnIlyLeuAlaGluPheGlyAlaGluPheIysAsnIle 240
b 709 AATTTGTGTTAAGACCTTTAACAAGTTGCTGAATTTGACAGACAGTTCAAGAACAT 768
y 241 GluThrAsnTrpThrArgIlePheLeuTrpSerGlyGluProThrTrpIleuGlyAsnGlu 260
b 769 GAAMCCAACTATACAGAAATATTTCTTACAGGAGAACCTTACTATCTGGAATGAA 828
y 261 ThrSerValPheGlyProThrGlyAsnIlySerThrIleuGlyLeuAlaIleIysArgPheTrp 280
b 829 ACATCTGTTTTGGGCCCAACAGAAACAGACTGTTGTTAGCCATTAATAAAGATTTAT 888
y 281 TyrProPheIysProIleuProThrIlyGluPheLeuSerIleuGlnIlePhe 300
b 889 TACCCCTTCAACACATTTGCCCACATAAGAAATTTGTTGAGTCTCTGCAAAATTTT 948
y 301 AspAlaValIleValIlyGlnPheTrpLeuPheTrpPheGluTrpPheLeu 320
b 949 GATGCAATGATGTCACAAACAGTTCTATTTGTTATATTTGAAATTTGTTT 1008
y 321 ProMetIysPheProPheIleIlyLeuThrTrpGlnGluIleProLeuProIleArgAsn 340
b 1009 CCTATGAATTTCCCTTTATTTAATAATACATATGAAATAAATCCCTTACCTACAGAAAC 1068
y 341 LysThrIleuSerGlyLeu 346
b 1069 AAAACACTCTCTGTTTA 1086

RESULT 4
15-09-397-945-26
Sequence 26, Application US/09397945
Publication No. US20030065139A1
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc. et al.
TITLE OR INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027P1
CURRENT APPLICATION NUMBER: US/09/397,945
CURRENT FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578

PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,581
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,577
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,563
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,313
PRIOR FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 470
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 26
LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1520)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-09-397-945-26

Alignment Scores:
Pred. No.: 3,736-230 Length: 1751
Score: 1916.00 Matches: 344
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 2
Query Match: 99.43% Indels: 0
DB: 13 Gaps: 0

US-10-010-050a-2 (1-346) x US-09-397-945-26 (1-1751)

Qy 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
Db 49 ATGCGCGGGGCGCGCGCGCGCTCGGGGACGCGGCTTCCGTGCTGCGGCGCTG 108
Qy 21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyIleProSerArgArgHis 40
Db 109 CTTTGGCTCGCGGTGTTCGGGCTGCTCCGGGCTCTCGGGCATCCCTCCGCGCCAC 168
Qy 41 TrpProValProIlyIysArgPheAspPheArgProIlyProAspProIlyCysGlnAla 60
Db 169 TGGCGGTGCTTCAACAGCCCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 228
Qy 61 LysTrpThrPheCysProThrGlySerProIleProValMetGluGlyAspAspAspIle 80
Db 229 AAGTATACCTTCTGTCACATGCTGCTACCTATCCAGTTATGAGGCTGATGATGACATT 288
Qy 81 GluValPheArgIleuGlnAlaProValTrpGluPheIlyTrpGlyAspLeuGlyHis 100
Db 289 GAAGTTTTCGATTACAGGCCCAAGTATGGAATTTAAATATGAGACCTCTCGGACAC 348
Qy 101 LeuIlyIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyIysAsnTrpTrp 120
Db 349 TTGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
Qy 121 MetGluTrpTrpGluLeuPheGlnLeuGlyAsnCysThrPheProIleuArgProGlu 140
Db 409 ATGAAATGTATGAACTTTTCCAACTTGCACTGTACATTTCCCATCTCCGACTGAA 468

101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyGlyAsnTyrThr 1..

PRIOR APPLICATION NUMBER: 0

```

1 APPLICANT: Chen, Weisheng
2 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
3 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
4 FILE REFERENCE: Aeomica-X-1
5 CURRENT APPLICATION NUMBER: US/09/864,761
6 CURRENT FILING DATE: 2001-05-23
7 PRIOR APPLICATION NUMBER: US 60/180,312
8 PRIOR FILING DATE: 2000-02-04
9 PRIOR APPLICATION NUMBER: US 60/207,456
10 PRIOR FILING DATE: 2000-05-26
11 PRIOR APPLICATION NUMBER: US 09/632,366
12 PRIOR FILING DATE: 2000-08-03
13 PRIOR APPLICATION NUMBER: GB 24263.6
14 PRIOR FILING DATE: 2000-10-04
15 PRIOR APPLICATION NUMBER: US 60/236,359

```

PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/609,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 22592
 LENGTH: 494

TYPE: DNA
 ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC001226.1
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
 OTHER INFORMATION: NT HIT: 915729771, EVALU0 0.00e+00
 OTHER INFORMATION: SWISSPROT HIT: 075503, EVALU0 2.00e-96
 OTHER INFORMATION: EST_HUMAN HIT: H01255.1, EVALU0 0.00e+00
 -09-864-761-22592

Alignment Scores:

Seq. No.: 2e-101 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Net Local Similarity: 100.00% Mismatches: 0
 Indels: 0
 Gaps: 0
 Indels: 0

:-10-010-050a-2 (1-346) X US-09-864-761-22592 (1-494)

184 AAlaVSTPValVlyGhAspAsngJlurhGlyIleYrYgLuThrTPrAsnVallys 203
 3 GCAAGTGGGAGAAACAGACATATAACAGAAATTTTATGACATGAGATGTAAAA 62
 204 AAlaSerProGluYsgLYAlaGluThrTPrPheAspSerTYrAspCyseSerlyPheVal 223
 63 GCACCCCAAGAAAGGAGGAGACATGCTGTGATTCCTACGACTGTCCAAATTTGTG 122
 224 LeuAlGThrPheAsnLyLeuAlaGluPheGlyAlaGluPheValAsnLlGluThrAsn 243
 123 TTAAGACCTTTAAACAGTTGGCTGAATTTGAGAGAGAGTTCAAGAAACCAAC 182
 244 TyrThrArgGlyPheLeuTySerGlyGluProThrTYrLeuGlyAsnGluThrSerVal 263
 183 TATACAGAAATTTCTTTACAGTGGAGACCTACTTATCTCGGAAATGAAACATCTGTT 242

QY 264 PheGlyProThrGlyAsnLySerThreGluYLeuAlaIleYsArgPheTYrYrProPhe 283
 DB 243 TTGGCCACAGAGAAACAGACTCTTGTTTACCATTAAGAAATTTATTAACCCCTTC 302
 QY 284 LysProHisLeuProThrLyGluPheLeuSerLeuGluLllePheAspAlaVal 303
 DB 303 AAACCACTTGGCAACTAAGAAATTTGTTGAGCTCTTGCAAAATTTGTATGAGTG 362
 QY 304 lleValHisLyGluPheTYrLeuPheTYrAsnPheGluTYrThrPheLeuPromelys 323
 DB 363 ATGTGCACAAACAGTTCTATTGTTTAAATTTGAAATATTGGTTTATTAACCTAAGAAA 422
 QY 324 PheProPheLleYsLleThrTYrGluGluIleProLeuProLlleArgAsnLySthLeu 343
 DB 423 TTCCTTTATTAAATTAACATTTGAAGAAATCTTTACTTACCTACGAAACAAACACTC 482
 QY 344 SerGlyLeu 346
 DB 483 TCTGTTTA 491

RESULT 8

US-10-040-739-825
 ; Sequence 825, Application US/10040739
 ; Publication No. US20020173635A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John

LaValle, Edward

Racie, Lisa

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS

NUMBER OF SEQUENCES: 1519

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/040,739

FILING DATE: 07-Jan-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/036,520

FILING DATE: 03-JUN-1998

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 825:

SEQUENCE CHARACTERISTICS:

LENGTH: 506 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 825:

US-10-040-739-825

Alignment Scores:

Seq. No.: 2,11e-81 Length: 506

Score: 727.00 Matches: 133

Percent Similarity: 100.00%
 at Local Similarity: 100.00%
 ery Match: 37.73%
 Gaps: 0

-10-010-050a-2 (1-346) x US-10-040-739-825 (1-506)

151 AAlaAlaCysPhePheGluGlyIleAspAspValHisTyrPheGluAsnGlyThrLeuVal 170
 99 GCTGCTCCTTTTGAAGGAATTCATGATGCTCACTGGAAGAAATGCGACATTTAGTT 158
 171 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaValTyrValGluAsp 190
 159 CAAGTAGCAACTATATACAGAAACATGTTCAACCAATGGCAAGTGGTGAACAGGAC 218
 191 AsnGluThrGlyIleTyrTyrGluThrTyrAsnValIleAspSerProGluIleGlyAla 210
 219 AATGAAACAGGATTTATTTATTCAGCATGGAATGTAAAGCCAGCCAGAAAGGGGCA 276
 211 GluThrTyrPheAspSerTyrAspCysSerIlePheValLeuValGlyThrPheAsnValLeu 230
 279 GAGACATGGTTTGAATTCCTTCAACAGCTGTTCAAAATTTGTTTAAAGGACCTTTAAACAAGTTG 338
 231 AlaGluPheGlyAlaGluPheAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 250
 339 GCTGAATTTGGAGCAGAGTTCAAGAACATAGAAACCACTATACAGATATTTCTTTAC 398
 251 SerGlyIleProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIle 270
 399 AGTGGGAACCTTACTTATCTGGGAAATGAAACATCTGTTTGGGCAACAGAAACAAG 458
 271 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPhe 283
 459 ACTCTTGTTTGGCCATTAAGAAATTTTATTAACCCCTTC 497

RESULT 9

8-09-864-761-5832

Sequence 5832, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263,6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 5832
 LENGTH: 474
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC00126.1
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
 OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 3.6
 OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3

US-09-864-761-5832

Alignment Scores:

Pred. No.: 4,24e-76
 Score: 684.50
 Percent Similarity: 86.67%
 Best Local Similarity: 86.00%
 Query Match: 35.52%
 DB: 9
 Gaps: 1

US-10-010-050a-2 (1-346) x US-09-864-761-5832 (1-474)

QY 153 CysPhePheGluGlyIleAspAspValHisTyrPheGluAsnGlyThrLeuValGlnVal 172
 Db 82 TGTGTTTTTAA----- 93
 QY 173 AlaThrIleSerGlyAsnMetPheAsnGlnMetAlaValTyrValGluAsnGlyThrLeuValGlnVal 192
 Db 94 -----CTGGAACATGTTCAACCAATGGCAAGTGGTGAACAGACATGAA 144
 QY 193 ThrGlyIleTyrTyrGluThrTyrAsnValIleAspSerProGluIleGlyThrLeuValGlnVal 212
 Db 145 ACAGGAATTTATTTACAGACATGGAATGTAAAGCCAGCCAGAAAGGGGCGAGAGAC 204
 QY 213 TyrPheAspSerTyrAspCysSerIlePheValLeuArgThrPheAsnIleGlyThrLeuValGlnVal 232
 Db 205 TGGTTGATTCCTACACAGCTGTTCCAAATTTGTTTGAAGACCTTTAAAGATTGGCTGAA 264
 QY 233 PheGlyAlaGluPheAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGly 252
 Db 265 TTTGGACAGAGTTCAAGAACATAGAAACCACTATACAAAGATATTTCTTAAACATGGA 324
 QY 253 GluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIleGlyThrLeuValGlnVal 272
 Db 325 GAACCTACTTATCTGGAAATGAAACATCTGTTTGGGCGAAGAGAAACAGACTCTT 384
 QY 273 GlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrIleGluPhe 292
 Db 385 GGTTCAGCCATTAAGATTTTATACCCCTTGAACACATTTGGCACTAAAGATTT 444
 QY 293 LeuLeuSerLeuLeuGlnIlePheAspAla 302
 Db 445 CTGTAGTCTCTTGAACATTTTGAATGCA 474

RESULT 10

10-264-049-1859
 Sequence 1859, Application US/10264049
 Publication No. US20040005579A1
 GENERAL INFORMATION:
 APPLICANT: Blisse et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 FILE REFERENCE: PA133P1
 CURRENT APPLICATION NUMBER: US/10/264,049
 CURRENT FILING DATE: 2002-10-04
 PRIOR APPLICATION NUMBER: PCT/US01/18569
 PRIOR FILING DATE: 2001-06-07
 PRIOR APPLICATION NUMBER: US 60/209,467
 PRIOR FILING DATE: 2000-06-07
 NUMBER OF SEQ ID NOS: 4360
 SOFTWARE: Patent In Ver. 3.1
 SEQ ID NO 1859
 LENGTH: 497
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (2)..(2)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (175)..(175)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (285)..(285)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (295)..(295)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (313)..(313)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (323)..(324)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (331)..(331)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (395)..(395)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (440)..(440)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (458)..(458)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (464)..(464)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (478)..(478)
 OTHER INFORMATION: n equals a,t,g, or c
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (483)..(483)
 OTHER INFORMATION: n equals a,t,g, or c
 3-10-264-049-1859

Alignment Scores:
 Pred. No.: 2.51e-49
 Score: 472.00
 Percent Similarity: 98.88%
 Best Local Similarity: 98.88%
 Query Match: 24.49%
 DB: 16
 Gaps: 0

US-10-010-050A-2 (1-346) x US-10-264-049-1859 (1-497)

QY 258 GYAENGUHTURSERVALPHEGIYPROTHRGIAENLYETHLEUGLYLEUALAILLEY 277
 DB 3 GGAATGAAACATCTGTTTGGCCCAAGGAAACAGACTCTGTTAGCCAAAA 62
 QY 278 ARGPHETTYTPROPHELYSPROHISLEUPROTHLYSGUAPHELEUSESERLEU 297
 DB 63 AGATTTTATACCCCTTCACACCACTTTGCCACTTAAGAAATTTCTGTGAGTCTTG 122
 QY 298 GUNILEPHEASPAIValIleValHisLYSGINPHEITYLEUPHETYSANPHEGIUTY 317
 DB 123 CAATTTTGTGATGCGATGATTTGCAACAAAGTTCTATTGTTATATNTGAATAT 182
 QY 318 TRPHELEUPROMELYSPHEPHEPHEILELYSILETHITYTGLUGIULEPROLEUPRO 337
 DB 183 TGGTTTTCACATGAAATTCCTTTATTAAATACATAGAGAAATCCCTTACCT 242
 QY 338 ILEARGANLYETHYLEUSERGTYLEU 346
 DB 243 ATCAGAAACAAACACTCTGTGTTTA 269

RESULT 11

US-09-867-701-2579
 ; Sequence 2579, Application US/09867701
 ; Patent No. US20020132237A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Agilate, Paul A.
 ; APPLICANT: Jones, Robert
 ; APPLICANT: Harlocker, Susan L.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.497
 ; CURRENT APPLICATION NUMBER: US/09/867,701
 ; CURRENT FILING DATE: 2001-05-29
 ; NUMBER OF SEQ. ID NOS: 10912
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 2579
 ; LENGTH: 351
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; US-09-867-701-2579

Alignment Scores:
 Pred. No.: 2.06e-35
 Score: 359.50
 Percent Similarity: 68.25%
 Best Local Similarity: 61.90%
 Query Match: 18.66%
 DB: 9
 Gaps: 3

US-10-010-050A-2 (1-346) x US-09-867-701-2579 (1-351)

QY 112 SERTHRLEUTHRGILYVAENLYTRHMEGLUTRPTRYGILEUPHEGINLEUGLYASN 131
 DB 3 AGTACATTACTGCGAAGACTTACACATGATGATGATGATGATGATGATGATGATGAT 62
 QY 132 CYSTHRPHEPROHISLEUARGPROGLUWELASPAIAPROPHETTPCYASNGINGLYALA 151
 DB 63 TGTACATTTCCCACTTCGCCACTGGAATGATGATGATGATGATGATGATGATGATGAT 122
 QY 152 ALACYSPEPHEGUGLYTLEASPAIYALHISTPLYSGLUANGLYTHLEUVALGIN 171
 DB 123 GCTGCTTTTGAAGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181

gment Scores: 0.00739 Length: 2667
 re: 112.00 Matches: 84
 cent Similarity: 32.17% Conservative: 36
 t Local Similarity: 22.52% Mismatches: 121
 ry Match: 5.81% Indels: 133
 Gaps: 19

-10-010-050A-2 (1-346) x US-10-198-846-9753 (1-2667)

20 LeuLeuTriPheAlaValAlProGlyTTPSer-----ArgValSer 33
 1790 CTGCTCTGGCTCAGC-----CCAGCTTGGAGTCTCCGCTGCTCACTCTCGGTCAGCG 1737
 34 GlyTleProSerArgArg-----HisTPProValProTyrLysArgPhePhe--- 50
 1736 GCCTTCTCCCGCAGCGCTTCCACACAC-CCTGTGCTCTTGAGACCCAGCCCTGTGTGA 1678
 51 ArgProLysProAspProTyrCysGlnAlaLysTyrPhe----- 64
 1677 AGTCCCTGGCCCGGCTCTCTTGACACACACCGAGCTTCCAGAACACGAGAGCACC 1618
 65 -----CysProthnGly----- 68
 1617 GGGAGCTGAGTGGAGACAGACCTCTATGCGTCCGCTCCCACTTCCCACTGAGTGAAT 1553
 69 -----SerProLysProValMetGluGlyAsp-----AspAspIle 80
 1557 GGGGACATGAGACAGCCCGCAGCTAGGACCATCTCTCACTGAGCTGGAGCCAGATC 1493
 81 GluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAspLeuLeuGlyHis 100
 1497 CAGCTATGTCTTCACTGCTCCG---TGGGATGCG---TGGCAT 1459
 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
 1458 CTCAGCCTTCTCCAGAGAGTGGACTCAGCTCAGCTCAGATGTCACATGTCACAA---ACG 1405
 121 MetGluTyrPyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGlu 140
 1404 CAAAGGTCTTCTGCTTACTTAATA---AACGTTCCCTTCCAGAACACAGCCCAA 1348
 141 -----MetAspAlaProPhe----- 145
 1347 CTCATCACTGTTCAATATGTCGCAAGTGAAGCTCTGCTCCCTTACCAGAGAGCAGC 1288
 146 ---TPCysAsnGlnGlyAlaAlaCysPhePheGluGly----- 157
 1287 GAGTGTGTCTGCAAGCAGAGAGTCTCTGAGTGTCTGAGAGATGGCCCTCAGAGCTCCAA 1228
 158 -----IleAspAspValHis---TyrLysGluAsnGlyThrLeuValGlnValAlaT 174
 1227 ACCCTCCAAATTCAGAGATGTCAGAGCGGCTCGGAGAGGGTCTCTGCTCTCAATCCAGC 1168
 174 IleLeuSerGlyAsnMetPheAsnGlnMetAlaLys-----TyrValL 188
 1167 GGGTCAAGCGGCTCTTGGCGGGGAGAAAGACCAACAGCCGCTTCCCTCTCTG3--- 112
 188 yseGlnAspAsnGlnThrGlyLysTyrTyrGluThrTyrPheValLysAsnProGluL 208
 1111 --TCACAGACAGCAGCAGGCTTCAGAGCAGAGCTGGGAGAAACATGTATGCCCCAT- 1055
 208 yseGlyAlaLeuThrTyrPheAspSerTyrAspCysSerLysPheValLysArgThrPhe 228
 1054 --GGTAAACCTGGGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1001
 228 snLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgLys 243
 1001 ----- 1001
 248 heLeuTyrSerGlyLysProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrG 263

Db 1000 -----GCTG 997
 Qy 268 LysnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuP 288
 Db 996 GACCCCTGCTCTTGGGCTGACCGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 937
 Qy 288 rothLysGluPhe-----LeuLeuSerLeuLeuGlnIlePheAspAlaValI 304
 Db 936 TTTCGACGAAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 877
 Qy 304 LevalHisLysGlnPheTyrLeuPheTyrAsnPhe 315
 Db 876 TCCTCTGACGAAATTTATTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 842
 RESULT 15
 US-10-369-493-33507
 ; Sequence 33507, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xiandeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ. ID NOS: 47374
 ; SEQ. ID NO 33507
 ; LENGTH: 3530
 ; TYPE: DNA
 ; ORGANISM: Desulfotobacterium hafnense
 US-10-369-493-33507
 Alignment Scores:
 Pred. No.: 0.0319 Length: 3530
 Score: 108.50 Matches: 74
 Percent Similarity: 38.80% Conservative: 42
 Best Local Similarity: 24.75% Mismatches: 114
 Query Match: 5.63% Indels: 71
 Db: 16 Gaps: 15
 US-10-010-050A-2 (1-346) x US-10-369-493-33507 (1-3530)
 Qy 5 AlaGlyAlaAlaArgGlyArgAlaSerTyrCys-----TyrAlaLeuAla 19
 Db 2224 GATCGCGCGGCAAGAGGCGGCACTATTATTATGACAGCAGACGACCAATCCCTGGAGACTTCA 2283
 Qy 20 LeuLeuTriPhe-----AlaValAlaProGlyTTPSerArgValSerGlyLysPro 36
 Db 2284 ATGGCTATATATGAGAAATGATTCATGGCCAGCGGCTGTCAAGTCTTCACTCACTATGAG 2343
 Qy 37 SerArgArgHisTyrProValProTyrLysArgPheArgPheArgProLysProAspPro 56
 Db 2344 AAAAAGCAAGAGCTCTTAAACGCAAAATGAAGACTTTGGCCCATGCGCAGCGCG 2403
 Qy 57 TyrCysGlnAlaLysTyrThrPheCysProThrGlyLysSerProLysProValMetGluGly 76
 Db 2404 CCAATTCCTATGTCATATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2457
 Qy 77 AspAspAspIleGluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAsp 96
 Db 2458 -----TTTATGTAAGCTTGGCCATTTGTGGAG-----GTCTCTTA- 2492
 Qy 97 LeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGly 116
 Db 2493 GTGATTGAGAGAGTACCAATCTG-----ACCATCACCGCG 2528
 Qy 117 LysAsn-----TyrThrMetGluTyrPyrGluLeuPheGlnLeuGlyAsn 131

```
2529 ACCAATGACATCAACCCCTGGGAACCATTCCTTTCTTCTGACGTTGCCGAGTTTATC 2583
      |||
132 CysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAla 151
      ::
2589 ATGCCCATTTGGCGACAGTGTCCACAGAGTTGAAATGCCATTATATATGCTTTGGCAGGGGCG 2643
      ::
152 AlaCysPhePheGlnGlyIleAspAspValHisTrpLysGluAsnGly-----ThrLeu 169
      |||
2649 GAACGGAATCTTCAACTCTCTGATGAAAGCCGGAAAGCATGAAAGGCAAGTACCTTG 2703
      |||
170 ValGlnValAlaThrIleSerGly---AsnMetPheAsnGlnMetAla---LysTrpVal 187
      |||
2709 GTCCATGTCAA-ACAGAGAAAGGGGCCCATATGTTGAACACGCTCGCATACCGATGGGTG 2767
      |||
188 LysGlnAspAsnGluThrGlyIleTyrTyrGluThrTrpAsn-ValLysAlaSerProG 207
      ::
2768 GGCTTGGAAATCTCTCAGGA-----AGATGGCATGTGTGACTATTCGGAGCT 2815
      |||
207 uLysGlyAlaGluThrTrpPheAsp-----SerTyrAspCysSerLysPheValIle 224
      |||
2816 TAAAGGAGAACTCCGTTCTTCTGATGTGACTTGGCTACGAGAGAAAGATGTTCT 2875
      |||
224 uArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTy 244
      |||
2876 GCAC----- 2879
      |||
244 rThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPh 264
      |||
2880 ---AACATCACTTGTATGCC---GAACCC-----GGGCAAAAGGTGGCTTTTGT 2923
      |||
264 eGlyProThr-----GlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
      |||
2924 GGGGGCTACGGGAGCGGCAAGACCACTATTACCAATCTCATCAACGTTTCTAT 2978
      |||
```

arch completed: April 25, 2004, 06:14:02
b time : 697.597 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame_plus.p2n model

on: April 24, 2004, 23:02:52 ; Search time 114.853 Seconds
(without alignments)
1671.819 Million cell updates/sec

le: US-10-010-050A-2
1937
1 MRBGGAGRGASMCWALAL.....IKITYEIPLPINRKTISGL 346

ring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

arched: 682709 seque, 277475446 residues
al number of hits satisfying chosen parameters: 1365418

imum DB seq length: 0
imum DB seq length: 200000000
it-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

mand line parameters:
DEL-frame+ p2n.model -DEV=xlh
/cgn2_1/USPTO.spool/US10010050/runat.22042004.113205.27569/app.query.fasta_...1372
3-issued Patents NA -QFMT=fastap -SUFFIX=rm -TINMATCH=0.1 -LOOFC=0
3-EXEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=trans40.cdi
IST=45 -DOCALLIN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
DB=LOCAL -OUTWMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
3ER=US10010050 @CEN 1.1.140 @runat.22042004.113205.27569 -NCPV=6 -ICPU=3
3 MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLBLOCK=100 -LONGIOG
3 TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
3AEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

tebase : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A COMB.seg:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seg:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seg:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seg:*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seg:*
6: /cgn2_6/ptodata/2/ina/backfile1.seg:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	Query Match	Length	DB ID	Description
1	94.5	4.9	14070	4	US-09-108-006C-2
2	92	4.8	1827	4	US-09-453-702B-22
3	91.5	4.7	1827	4	US-09-489-039A-6332
4	91.5	4.7	7718	4	US-09-976-594-244
5	90.5	4.7	1349	1	US-08-290-448A-73
6	90.5	4.7	1349	1	US-08-290-448A-73
7	90.5	4.7	1349	1	US-08-175-069A-73
8	90.5	4.7	1349	4	US-08-461-939B-73
9	90.5	4.7	1349	4	US-08-461-939B-73
10	90.5	4.7	6418	1	US-08-480-528A-11
11	90.5	4.7	6418	1	US-08-479-666-11
12	90.5	4.7	6418	5	PCT-US93-10520-11

13	90	4.7	3113	4	US-10-028-272-1
14	89.5	4.6	1328	1	US-08-290-448A-58
15	89.5	4.6	1328	1	US-08-290-448A-58
16	89.5	4.6	1328	1	US-08-175-069A-58
17	89.5	4.6	1328	4	US-08-461-939B-58
18	89.5	4.6	1328	4	US-08-461-939B-58
19	89	4.6	12394	4	US-09-488-856A-10
20	88.5	4.6	4440	3	US-07-792-600-1
21	88.5	4.6	4440	3	US-09-157-021-1
22	88.5	4.6	4440	3	US-09-156-842-1
23	88.5	4.6	4440	3	US-09-591-514-1
24	88.5	4.6	5433	3	US-09-157-021-35
25	88.5	4.6	5433	3	US-09-156-842-35
26	88.5	4.6	5433	4	US-09-152-655-1363
27	88.5	4.6	5433	4	US-09-591-514-35
28	88.5	4.6	5433	4	US-09-620-312D-253
29	87.5	4.5	537	4	US-09-134-000C-219
30	86.5	4.5	13930	4	US-09-976-594-1011
31	86.5	4.5	13930	4	US-09-976-594-1011
32	86	4.5	429	4	US-09-252-991A-6217
33	86	4.5	1398	4	US-09-252-991A-7174
34	86	4.5	1398	4	US-09-252-991A-6413
35	86	4.5	1575	4	US-09-252-991A-7330
36	86	4.5	1890	4	US-09-252-991A-6330
37	86	4.5	2166	3	US-09-232-197-101
38	86	4.5	2166	4	US-09-232-197-101
39	86	4.5	2166	4	US-09-232-201-101
40	86	4.5	2166	4	US-09-232-195-101
41	86	4.5	3577	4	US-09-620-312D-457
42	85.5	4.4	438	4	US-09-252-991A-9484
43	85.5	4.4	1533	4	US-09-252-991A-9378
44	85.5	4.4	2043	4	US-09-252-991A-9459
45	85.5	4.4	2274	4	US-09-252-991A-9347

ALIGNMENTS

RESULT 1
US-09-108-006C-2
Sequence 2, Application US/09108006C
Patent No. 6524613
GENERAL INFORMATION:
APPLICANT: Steer, Clifford J.
Kren, Betsy T.
Bandyopadhyay, Paramita
Roy-Chowdhury, Tanya
TITLE OF INVENTION: Hepatocellular Chimeraplasty
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kimeragen, Inc.
STREET: 300 Pleasant Run
CITY: Newtown
STATE: PA
COUNTRY: USA
ZIP: 18940
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/108,006C
FILING DATE: 30-Jun-1992
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/054,288
FILING DATE: 30-APR-1997
APPLICATION NUMBER: 60/054,837
FILING DATE: 05-AUG-1997
APPLICATION NUMBER: 60/064,996
FILING DATE: 10-NOV-1997
APPLICATION NUMBER: 60/074,497
FILING DATE: 12-FEB-1998

```
QY      249 uUySerGIgLuProTHrTYRLeugLYAsnGLuThreValPheGlyProTHnIGLYAs 269
           |||::: :::::
Db      12549 TCAGGAAGCCCAAGCCAGTTTCCACGGAGCTCAGAAGATAACGGITTAGTGCGTTGGTACG 12608

QY      269 nIysThrLeugLYeAlaAlleLyS----- 277
           |||||:::||||
Db      12609 AGTTACTACAAGATTTCATATATAAAGGTCAAGCATCTGATTAACCTCACATTGATTTCCT 12668

QY      278 -----ArgPhetYrTyRPheProHIsleUProTHrLYSGluPheLeu 294
           |||::: :::::
Db      12669 GAACCTGCCCAAGTTTCAGTTTTCCGGGAAACCTGGGATATACACTAAGGAGAACCTTGG 12728

QY      294 uSeRLeuLengInIllePheSepAlaValleValHisLYSGlnPheTYr----- 310
           |||::: :::::
Db      12729 CACTATGTTCC---ATTAAGGAGAGGTAGGAGACGGTACTGTTCCCAGTATATTCAGAAAGTCCA 12785

QY      311 -----LeuPheTyraSPHeGuTYrtTpheLeuProMeILysPhePr 325
           |||||:::||||
Db      12786 TAATGTTGCAGAAAATACTGTTTTCCTATTTCCAAAGCCTAGTGAATTACACTTCCTTCCA 12843

QY      325 oPhelleyS 328
           |||
Db      12846 GTTAAGGAAA 12855

RESULT 2
US-09-453-702B-22
; Sequence 22, Application US/09453702B
; Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
          Burland, Valerie
          Perna, Nicole T.
          Plunkett, Guy
          Welch, Rod
TITLE OF INVENTION: NO. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSER: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 4643
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-453-702B-22
```

```

ed. No.: 3.63 Length: 4643
Cre: 93.00 Matches: 59
Percent Similarity: 31.38% Conservative: 32
Percent Local Similarity: 20.34% Mismatches: 93
Query Match: 4.83% Indels: 106
Gaps: 15

1-10-010-050A-2 (1-346) x US-09-489-039A-6332 (1-1827)

1 MetArgValGlyAlaValAlaValArgAlaSerTyrCysTyrAlaValAlaValLeu 20
3363 CTGCGCGCGGCAAGGCGCAATAGTGT----- 3392
21 LeuTyrLeuAlaValAlaValProGlyTyrSerArgValSerGlyIleProSerArgHis 40
3393 -----AACTGATCTGACACAGCGGAGGATCTCTGACTG 3431
41 TrpProValProTyrIleValArgPheArgPheArgProIleProArgProTyrCysGlnIle 60
3432 CAGCGCGGAGCCA-----GACCCGATGCGGATCCAGAGCCAAACCCGAAATCCA 3479
61 LysTyrThrPheCysProThrGlySerProIleProValMetGluIleAspAspIle 80
3480 GAGCCGACCCCTAACCCGACCTACGCGCGGCTCCGATCTGAATGTGATATACCTG 3539
81 GluValPheArgLeuGlnAlaProValTyrGluPheIleTyrGlyAspLeuGlnHis 100
3540 -----CGACCGGAGGCG-----GGTAGCTACATTGGGAAAC 3559
101 LeuIleIle-----MetHisAlaIleGly----- 109
3570 CTGGACGACGAGCAATACATGTTACACACGCTGTGATGAGCTGGGTAAATACGTAAC 3629
110 PheArgSerThrIleThrGly-----LysAsnTyrThrMet----- 121
3630 TAAACCCAGATGGGAGGAGGAGGAGCAAAACCACTATGTGATGCCCATGAAGT 3689
122 -----GluTyrTyrGlu-----Leu 126
3690 GGTCAATATAATGGCTGATGGACGCGGCGGACGTAACCCAAAGCAATCGCTATGT 3749
127 PheGlnLeuGlyAsnGlyThrPheProHisLeuArgProGluMetLarPalaProPheTyr 146
3750 CTGCAACTGGGAGGCG-----GATGTGCGGAGTGG 3779
147 CysAsnGlnGlyAlaIleAlaCysPhePheGluGlyIleAspAspValHisTyrIleGluAsn 166
3780 AGCCAAACGCGACGACGACCGCTGCGATGTGGGTGATGCGCGGATATGGAACAGCGAC 3839
167 GlyThrIleValGln-----ValAlaThrIleSerGlyAsnMet 179
3840 AGCAAAACCAATTTCTCGCGCAACCGGTTATGTCGAAAGAGGTGTAACGATATAGC 3899
180 PheAsnGlnMetAlaIleTyrValIleGlnAsp-----AsnGluThrGlyIleTyrTyrGlu 198
3900 ACAAAGCTCTATGCACTGATGCGATGCGATGCGATGCGGATGCGGATATCTGCAC 3959
199 ThrTyrAsnValIleValAlaSerProGluIleGlyAlaGluThrTyrPheAsp----- 215
3960 AGTTGG-----GGCAGTACAGCTGGTTTGATTAACACAGT 3995
216 -----SerTyrAspCysSerTyrPheValIleuArgThr----- 226
3996 AAAAGGATGACTTACAAAGTGAATCTTAAATACAAAGATTAACCGCTTCACTGAA 4055
227 -----PheAsnIleValAlaGluPhe 233
4056 GCTGATACAAACAAATTAAGTGAATTT 4085

```

```

GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OR INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 6332
LENGTH: 1827
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6332

Alignment Scores:
Pred. No.: 1.04 Length: 1827
Score: 92.00 Matches: 77
Percent Similarity: 31.53% Conservative: 34
Best Local Similarity: 21.88% Mismatches: 100
Query Match: 4.77% Indels: 141
Gaps: 23

US-10-010-050A-2 (1-346) x US-09-489-039A-6332 (1-1827)

29 TrpSerArgValSerGlyIleProSerArgHisTyrProVal----- 43
678 TGGCGGAGCGGCGGCGCTGCTCCGCTGCGCCACACACCGCGCGGAAATGCGCAT 737
44 -----ProTyrIleValArgPheAspPhe----- 50
738 GATCTCGACGCGGTGTTCACCAATACCGGAGTCCCATCTCCGTGACCGGTACCA 797
51 -----ArgProIleProAspProTyrCysGlnAlaIleTyrThrPhe 64
798 GCAGGACGCGGCGCGCGCGGCGGACGATCCGACCTCCGCGGAGCTGTTACTT 857
64 eCysProThrGlySerProIleProValMetGluIleAspAspIleGluValPheArg 84
858 CTCAGAGGAGGCG----- 870
84 GluGlnAlaProValTyrGluPheIleTyrGlyAspLeuGlnHisIleuIle 104
871 -----CAGGCGCACACTCG-----CTGGCTATGCGCACTCG-- 903
104 HisAspAlaIleGlyPheArgSerThr-----LeuThrGlyLysAsnTyrThrMetGlu-- 122
904 -----CGAAGCTGACATATGCTGAGACGCGCTGTCACAGATCTATGCGGTGAAGA 959
123 -----TyrTyrGluLeuPheGlnLeuGlyAsnGlyThrPheProHisIle 137
960 CAGTATGTCAGCGCACTGG----- 978
137 uArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaIleCysPhePheGlnIle 157
979 -----CTGAAGGCGCA-----TGAAGCATGACCGCTGCGCT-- 1011
157 YIleAspAspValHisTyrIleGluAsnGly-----ThrLeuValGlnIle 172
1012 -CTTGAAGCTGTCATATGCTGAGGAGGCGGCGGCGGCGGAAATTAACCTGACAGATAT 1070
172 IAlaThrIleSer-----GlyAsnMet 179
1071 CGCCGCAATTAACCCAGGACGAGCAAGCAAGCCAGCGGAGCTTTGTCTTGGGAAACA 1130
179 tPheAsnGlnMetAlaIleTyrValIleGlnAsp-----AsnGluThr 193
1131 TTTTGGCAGCGGCGGCGGCGGCTGTCAGGCGGATCCGAAAGCGCGGAGTAAGTATGCG 1190
193 tGlyIleTyrTyrGluThrTyr-----AsnValIleAlaSerProGlu 207
1191 TGGCTTCACTTCCGATCTGGGATTCCTGCGCAATACCGATATCTCTTAAGATCCGCA 1250

```

RESULT 3
 US-09-489-039A-6332
 ; Sequence 6332, Application US/09489039A
 ; Patent No. 6610836

2y 207 ULys---GlyValGluThr-----ThrPheAspSerTyrAspSer----- 220
 1251 GAAATATGACCCCGACCTGTATGGCGTGATGATATATATGCGCCGCGGTGTGCA 1310
 2y 221 -LysPheValLeuArgThrPheAsnLeu-----AlaGluPhe----- 233
 1311 TCAGACAGAGCTGGCGAGCTTCAACACAGCTGCACGCCATGATACGGCGCTTTAATC 1370
 2y 234 -----GlyAlaGluPheLysAsnLeuThrAsnTyrThrArgIlePheLeuTyrSe 251
 1371 TCTTCTTGCAAAAGCTAGCGCGCTGCGCTGGCGGTGGTGTGCTGTTCAGCTGGCC 1430
 2y 251 rGlyGluPro---ThrTyrLeuGlyAsnGluThrSerValPheGly-----Proth 267
 1431 GGGGGGTACCGTCATCTACTATGGCGAGGTGGCGGTGATGCAATACATCCGTT 1490
 2y 267 rGlyAsnTyrThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisle 287
 1491 CTGCCGCAACCG-----TTCCCGTGGATCCGGCGCT 1523
 2y 287 uProThrLysGluPheLeuSerLeuGln 298
 1524 GCAGATACCCAG---CTGCTGGCGCTGTATCAG 1554
 Db

RESULT 4
 US-09-976-594-244
 ; Sequence 244, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LAYER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; PRIOR FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 244
 ; LENGTH: 7718
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 322303.15
 ; NAME/KEY: unsure
 ; LOCATION: 6339-6362, 6967-7026
 ; OTHER INFORMATION: a, t, c, g, or other
 US-09-976-594-244

Alignment Scores:
 Pred. No.: 12.6 Length: 7718
 Score: 91.50 Matches: 71
 Percent Similarity: 31.40% Conservative: 37
 Best Local Similarity: 20.64% Mismatches: 133
 Query Match: 4.75% Indels: 103
 DB: 4 Gaps: 15

US-10-010-050a-2 (1-346) x US-09-976-594-244 (1-7718)

QY 41 TrpProValProTyrLysArgPheAsp-----PheArgProLysProAspProTyrLys 58
 Db 3443 TGGATGGCTCCGAATCTATCTTTGACAAATCTACAGCACAGACGCGTGTGCT 3502
 QY 59 GlnAlaLysTyr-----ThrPheCysProThrGlySerProIleProValMetGlu 75
 Db 3503 TAGGAGATATGGCTGGGAAATCTTCTCTTAGTGGGTCTCCATACCCAGAGATACAA 3562
 QY 76 GlyAspAspAspIle----- 60
 Db 3563 ATGGATGAGGACTTTTGACCTGCTGAGGAAAGCATGAGATGAGAGCTCTGAGTAC 3622

QY 81 -----GlyValPheArgLeuGlnAlaProValTrpGlu----- 91
 Db 3623 TCTACTCTGAATATCTATCTGATCATGCTGAGCTGCACAGACCCAAAGAAAG 3682
 QY 92 -----PheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMet 104
 Db 3683 CCAAGATTTCAGAACTTGAGAAACCTAGGTGATTTGCT---CAAGCAATGTACAA 3739
 QY 105 HisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyr 124
 Db 3740 CAGAT-----GTYAAAGCTACATCCCA-----ATC 3766
 QY 125 GluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGluMetAspAlaPro 144
 Db 3767 AATCCCATCTGACAGAAATAGTGTTTACATCACTCACTGCTCTCTGAGAGAC 3826
 QY 145 PheTyrCysAsnGlnGlyAlaIleCysPhePheGlu-----GlyIleAspAspValHis 162
 Db 3827 TTCTTCAAGAAAGTATATTCAGCTCCGAAGTTTATTCAGAGACCTTGATGATGTGCA 3886
 QY 163 TrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln 182
 Db 3887 TATGTAAATGCTTCAAGTTTATGAGCTGGAAGAAATC-----AAACCTTGAAGAA 3940
 QY 183 MetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyr-----TyrGlu 198
 Db 3941 CTTTTA-----CCGAATGCCACTCTCCATGTTGATGACTACACAGGCGAC 3985
 QY 199 ThrTrpAsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAsp 218
 Db 3986 AGCAGCACTGTGGCTCTCCATGCTGAGGCGCTTCACCTGAGTCACTGAGCAAAACC 4045
 QY 219 CysSer-----LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGly----- 234
 Db 4046 AAGGCTGCTCAAGATTGACTTGAAGTACACAGTAAAGTAAAGATCGGGGCTGTCT 4105
 QY 235 -----AlaGluPheLysAsn 239
 Db 4106 GATGTACAGAGCCCAAGTTTCTGCCATTCAGCTGTGGGACAGTCAGGAAAGGCAAGGC 4165
 QY 240 IleGluThr----- 242
 Db 4166 AGGTTCACTACGACACAGCTGAGTGAAGAAAGAAATCGGCTGCTCCCGCCCA 4225
 QY 243 AsnTyrThrArgIlePheLeuTyrSerGlyGluPro-----ThrTyr 256
 Db 4226 GACTACAACTGGGTGCTCTGTACTCCACCCCATCTAGATTGACAGAGACCTT 4285
 QY 257 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle 276
 Db 4286 ATTCTAGAGACACATGTGTAATTATATACCCCGAGAAACTAGCTTTGGCAGATATTATGC 4345
 QY 277 LysArgPheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuSerLeu 296
 Db 4346 ATATATTAAGTTTAAACCTTATCTTCCATGAGGACGACGCTCTTTGTATTTT 4405
 QY 297 LeuGlnIlePhe 300
 Db 4406 ATAGTGCTTTT 4417

RESULT 5
 US-08-290-448A-73
 ; Sequence 73, Application US/08290448A
 ; Patent No. 5676954
 ; GENERAL INFORMATION:
 ; APPLICANT: Rogers, Bruce
 ; APPLICANT: Klapper, David G.
 ; APPLICANT: Rafnar, Thorunn
 ; APPLICANT: Kuo, Mei-chang
 ; TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
 ; NUMBER OF SEQUENCES: 93
 ; CORRESPONDENCE ADDRESS:

```

ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEO ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-290-448A-73

Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 4.70% Indels: 93
DB: 1 Gaps: 16

US-10-010-050A-2 (1-346) X US-08-290-448A-73 (1-1349)
QY 65 CysPro-----ThgIYserProIleProValMetGIuGIYAsp 77
DB 526 TGTCACAGAGCGCATGATTAGTCCACAGATGTCACCAATTTTAAAGACAAAGTGAT 585
QY 78 AspAspIleGIuValPheArgLeuGIuAlaProValTrpGIuPheIYsTYrGIYAspLeu 97
DB 586 GGTGATCTTAATATTGTCGTGGTAGTTCACAAATATCG----- 624
QY 98 LeuGIYHisLeuYsIleMetHisAspAlaIleGIYPheArgSerThrIleuThrGIYlys 117
DB 625 ATGCACCATGTGCTGCTCAGTAAGGCTCCGATGGCTGCATATCACCTCGGAGC 684
QY 118 AsnTYrThrMetGIuTrpTYrGIuLeuPheGIuLeuGIYAsnCYserThrPheProHisLeu 137
DB 685 TCACACGTG-----ACCGTTTCCACCTGCAAAATTC----- 714
QY 138 ArgProIleuMetAspAlaProPheTrpCysAsnGIuAlaIaIaCysPhePheGIuGIY 157
DB 715 -----ACCCAAACCAATTTGATATTATTCCTGGG 744
QY 158 ILAspAspValHisTrpIYsGIuAsnGIYThrIleuValGIuValAlaThrIleSerGIY 177
DB 745 GCGTATGACACCCCTATTCAGAAATAAAGCATGCTA-----GCAAGGTGAGCATTC 795
QY 178 AsnMetPheAsnGIuMetAlaIaIaYsTrpValIYsIleAspAsnGIuThrGIYIleTYr--- 196

```

```

DB 796 AAGATGTTCCACCGATCAGCTTGACCAAGAAATGCCTAGATGTAGATTGGGTTTTCGA 855
QY 197 -----TYrGIuThrTYrAsnValIYsAlaSerProGIuIYsGIYAlaGIu 211
DB 856 GTGTTAAACAACACTGACAGATGGGAAAGTAAGCATGCTGTGTAAGCTCGGCCCA 915
QY 212 Thr-----TrpPheAspSerTYrAspCysSerIYsPheValIeu 224
DB 916 ACTATACTCGACCAAGGAAAGACAGATTCCTCGCCCCCATGATATCATCAAG----- 966
QY 225 ArgThrPheAsnIYsLeuAlaGIuPheGIY-----AlaGIuPheIYsAsnIleGIu 241
DB 967 -----AAAAATGCTTTCAGGACGACTGTACTGGCAACGACAGATGCTGTGAAC 1020
QY 242 ThrAsnTYrThrArgIlePheLeuTYrSerGIYGIuProThrTYrIleuGIYAsnGIuThr 261
DB 1021 TCGAGAACAGATAGAGACTGCTGTGAATATGCT----- 1053
QY 262 SerValPheGIYProThrGIY---AsnIYsThrLeuGIYLeuAlaIleIYsArgPheTYr 280
DB 1054 GGTATTTTCTCCATCCGGGCTGTGATCCAGTCTAACCCCTGAGCAAAAGCAGGATG 1113
QY 281 TYrProPheIYsProHisLeuProThrIYsGIuPheLeuSerIeuLeuGIuIlePhe 300
DB 1114 ATTCACGCTGACCA-----GGAGAAAGCGTTCTTAGACTCTAGTAGTGCT 1161
QY 301 AspAlaValIleValHisIYs-----GlnPheTYrLeuPhe 312
DB 1162 GGTGATCTCTCATGCGCATACAGAGACCTTGCTAGACCTGGCCAAATTCCTAGCTTY 1221
QY 313 TYrAsn-----PheGIuTYrTYrPheLeuProMetIYsPheProHelleIYsIle 329
DB 1222 TATTAATATCATTAATATTATTATTATTATT-----TTGTATATT 1263
QY 330 ThrTYrGIuGIuIleProLeuProIleArgAsnIYsThrIeu 343
DB 1264 TTAATATGA-----CATTAAGTTCAAGTACTCTA 1293

RESULT 6
US-08-290-448A-73
Sequence 73, Application US/08290448A
Patent No. 5698204
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207

```

```

REFERENCE/DOCKET NUMBER: IMI-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-290-448A-73

Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 4.70% Indels: 93
DB: 1 Gaps: 16

US-10-010-050A-2 (1-346) x US-08-290-448A-73 (1-1349)
QY 65 CysPro-----ThGlySerProIleProValMetGluGlyAsp 77
DB 526 TGTCAGAGAGGAGGATTAAGTCCACAGATGCTCCACCAATTTTAAAGCAACAAAGTCAT 585
QY 78 AspAspIleGluValPheArgIleuGluAlaProValTrrpGluPheLysTyrGlyAspLeu 97
DB 586 GGTATGCTATTAATAGTTGCTGCTAGTCTACAAATATG----- 624
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
DB 625 ATCGACCATTTGCTCCCTCAGTAAGGCTTCGATGGGCTGCTGATATCACCCCTGGCAGC 684
QY 118 AsnTyrThrMetGluTrrpTyrGluLeuPheGluGluGlyAsnCysThrPheProHisLeu 137
DB 665 TCACACGGG-----ACCGTTTCCAACTGCATAATTC----- 714
QY 138 ArgProGluMetAspAlaProPheTrrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
DB 715 -----ACCCAAACCAATTTGATTTATGCTCGGG 744
QY 158 IleAspAspValHisTrrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
DB 745 GGTGATGACACCCATTATCAAGATTAAGGCATGCTA-----GCAACGGTACATTC 795
QY 178 AsnMetPheAsnGlnMetAlaLysTrrpValLysGlnAspAsnGluThrGlyIleTyr--- 196
DB 796 AACATGTTCAACCCATGACCTTGACCAAGAAATGCTTAATGATTTGGGTTTTCGA 855
QY 197 -----TyrGluThrTrrpAsnValLysAlaSerProGluLysGlyAlaGlu 211
DB 856 GTGCTTAACAACAATACTAGACAGATGGGGAACGTACGCATCGGTGATAGCTGGGCCCA 915
QY 212 Thr-----TrrpPheAspSerTyrAspCysSerLysPheValLeu 224
DB 916 ACTATATCTCAGCCAAAGGAACAGATTCTTCCGCCCGCATGATATCATCAAG----- 966
QY 225 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 241
DB 967 -----AAAAATGCTTACCGAGAGACTGCTACTGGCAAGCAGAGATGCTCGGGAAC 1020
QY 242 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrIleuGlyAsnGluThr 261
DB 1021 TCGAAGAACGATAGAGACTTGTGAATGGT----- 1053
QY 262 SerValPheGlyProThrGly---AsnLysThrIleuGlyLeuAlaIleLysArgPheTyr 280
DB 1054 GCTATTTTCTCCATCCGCGGTCTGATTCAGTGTACACCCCTGAGCAAAACAGGAGATG 1113

```

```

QY 281 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuLeuGlnIlePhe 300
DB 1114 ATTCGAGCTGAACCA-----GGAGAAAGCCGTTCTAAGACTCACTAGTGTCT 1161
QY 301 AspAlaValIleValHisLys-----GlnPheTyrLeuPhe 312
DB 1162 GGTGTACTCTCATGACATCAAGAGACCTTGCTAAGACCTGGCCAAATCTTAGCTTT 1221
QY 313 TyrAsn-----PheGluTyrTrrpPheLeuProMetLysPheProPheIleLysIle 329
DB 1222 TATTAATATCATTAATAACTTATTATTATTATT-----TTGATATT 1263
QY 330 ThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 343
DB 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293

RESULT 7
US-08-175-069A-73
Sequence 73, Application US/08175069A
Patent No. 5776761
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Ratnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHYE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175, 069A
FILING DATE: December 29, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529, 951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325, 365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-175-069A-73

Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 4.70% Indels: 93

```

```

DB: 1 Gaps: 16
US-10-010-050a-2 (1-346) x US-08-175-069A-73 (1-1349)
QY 65 CysPro-----ThrsYserProIleProValMetGluGlyAsp 77
Db 526 TGTCGAGGAGCATGATTAAAGTCCAGATGGTCCACCAATTAAAGACAAAGATGAT 585
QY 78 AspAspIleGluValPheArgLeuGlnAlaProValTrrpGluPheLysTyrGlyAspLeu 97
Db 586 GGTGATGCTATAAATCTTCTGTTAGTTCAAAATATG----- 624
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
Db 625 ATGCACCATGCTCGCTCAGTAAGGCTCCGATGGCTCTCGATATTCACCTCGGCAGC 684
QY 118 AsnTyrThrMetGluTrrpTyrGluLeuPheGlnLeuGlyAsnCySerThrPheProHisLeu 137
Db 685 TCACAGCTG-----ACCGTTCCACTGCAAAATTC----- 714
QY 138 ArgProGluMetAspAlaProPheTrrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
Db 715 -----ACCCACACCAATTGATATTATGCTCGGG 744
QY 158 IleAspAspValHisTrrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
Db 745 GCTGATGACACCCATTATCAAGATAAAGCATGCTA-----GCCACGGTAGCATTC 795
QY 178 AsnMetPheAsnGlnMetAlaLysTrrpValLysGlnAspAsnGluThrGlyIleTyr--- 196
Db 796 AACATGTTCCAGCATTCACGTTGACCAAAGATGCTGATGATGATTTGGGTTTCCAA 855
QY 197 -----TrrpGluTrrpAsnValLysAsnSerProGluLysGlyAlaGlu 911
Db 856 GTCGTTAACAAACACTACAGACATGGGGAACGATGCCGCTGCTGCTCGGCCCA 915
QY 212 Thr-----TrrpPheAspSerTrrpCysSerLysPheValLeu 924
Db 916 ACTATACTACGCCAAGGAAACAGATCTCTGCCCCGATGATATCATCAAG----- 966
QY 225 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 941
Db 967 -----AAAATGCTTAGGAGGACTGTTACTGGCAACGACGATGATGCTGTGAAC 1020
QY 242 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrIleGlnLysAsnGluThr 261
Db 1021 TGGAGAACAGATAGAGCTGCTGAAATATGCT----- 1053
QY 262 SerValPheGlyProThrGly---AsnLysThrIleGlnLysAlaIleLysArgPheTyr 280
Db 1054 GCTATTTTCTCCATCCGCTCGATCCAGTGCCTAACCCCTGAGCAAAAAGCAGGATG 1113
QY 281 TyrThrPheLysProHisLeuProThrLysGlnPheLeuLysSerLeuLeuGlnIlePhe 300
Db 1114 ATTCACAGTGAACCA-----GGAGAACCCCTTCTTAAGCTCACAGTAGAGTCT 1161
QY 301 AspAlaValIleValHisLys-----GlnPheTyrLeuPhe 312
Db 1162 GGTGATGCTCTCATGCTACCAAGAGACCTGTTAAGACCTGGCCAAATCTCTAAGCTTT 1221
QY 313 TyrAsn-----PheGluTyrTrrpPheLeuProMetLysPheProPheIleLysIle 329
Db 1222 TATATATATCATATAATCTTATTTTATTTTAT-----TTTGATATTT 1263
QY 330 ThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 343
Db 1264 TTATATGAA-----CAATTACGTTCAAGTACTCTA 1293

```

RESULT 8
 US-08-461-939B-73
 ; Sequence 73, Application US/08461939B
 ; Patent No. 6335019
 ; GENERAL INFORMATION:

```

APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Ratner, Thorum
APPLICANT: Kuo, Mei-Chang
TITLE OF INVENTION: Methods For Treating Sensitivity To A
TITLE OF INVENTION: Protein Allergen Using Peptides which Include A T Cell Epitope
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,939B
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/464,000
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CNDV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
FAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-461-939B-73
Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservatve: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 4.70% Indels: 93
DB: Gaps: 16
US-10-010-050a-2 (1-346) x US-08-461-939B-73 (1-1349)
QY 65 CysPro-----ThrsYserProIleProValMetGluGlyAsp 77
Db 526 TGTCGAGGAGCATGATTAAAGTCCAGATGGTCCACCAATTAAAGACAAAGATGAT 585
QY 78 AspAspIleGluValPheArgLeuGlnAlaProValTrrpGluPheLysTyrGlyAspLeu 97
Db 586 GGTGATGCTATAAATCTTCTGTTAGTTCAAAATATG----- 624
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
Db 625 ATGCACCATGCTCGCTCAGTAAGGCTCCGATGGCTCTCGATATTCACCTCGGCAGC 684
QY 118 AsnTyrThrMetGluTrrpTyrGluLeuPheGlnLeuGlyAsnCySerThrPheProHisLeu 137

```

```

Db      685 TCACACGTG-----ACCGTTTCCAACTGCAGAAATTC----- 714
Qy      138 ArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
Db      715 -----ACCCAAACACCAATTTGATTATTGCTGGG 744
Qy      158 IleAspAspValHisTrrPylsGluAsnGlyThrlleuValGlnValAlaThrIleSergly 177
Db      745 GCTGATGACACCCATTATCAAGATTAAGGAGATGTA-----GCAACGGTAGCATTC 795
Qy      178 AsnMetPheAsnGlnMetAlaLysTrrValIleGlnAspAsnGluThrGlyIleTyr--- 96
Db      796 AACATGTTCAACCATCACTGACGTTGACCAAGATGCTGATGATTTGGGTTTCCAA 855
Qy      197 -----TyrGluThrTrrAsnValIleAlaSerProGluIlySglValGlu 211
Db      856 GTCGTTAACAAACAACACTACGACAGATGGGAAACGTACCGTGGTAGCTGGCCCA 915
Qy      212 Thr-----TrrPheAspSerTyrAspCysSerIysPheValIleu 224
Db      916 ACTATACTCAAGCCCAAGGAAACAGATTTTTCGCCCCCATGATATCATCAAG----- 366
Qy      225 ArgThrPheAsnIlySleuAlaGluPheGly-----AlaGluPheIysAsnIleGlu 241
Db      967 -----AAAATGCTTACGAGGAGACTGTAAGTCAACGAGAGAGTGCATGTCGTGAAC 1020
Qy      242 ThrAsnTyrThrArgIlePheLeuTyrSergIlyGluProThrTyrTrleuGlyAsnGluThr 261
Db      1021 TGGAGAACAGATAGAGACTTGTGAAATGCT----- 1053
Qy      262 SerValPheGlyProThnGly---AsnIlySthrleuGlyLeuAlaIleIysArgPheTyr 280
Db      1054 GCTATTTTCTCCATCGGCTGTGATCCAGTCTTAACCCCTGGACAAAAGCAGGATG 1113
Qy      281 TyrProPheIysProHisIleuProThrIlySglPheIleuSerIleuGlnIlePhe 300
Db      1114 ATTCACAGCTGAACCA-----CGAGAGACCGTTTAAAGACTCACTAGTAGTGCT 1161
Qy      301 AspAlaValIleValHisIys-----GlnPheTyrLeuPhe 312
Db      1162 GGTGTAATCTGTCATGCAAGAGACCTTGTAGACCTGGCCCATTTCTTAAGCTTT 1221
Qy      313 TyrAsn-----PheGluTyrTrrPheLeuProMetIysPheProPheIleIysIle 329
Db      1222 TATATTAATCATTAATCTTATTATTATTATT-----TTTGATATT 1263
Qy      330 ThrTyrGluGluIleProLeuProIleArgAsnIlySthrIleu 343
Db      1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293

RESULT 9
US-08-464-000-73
; Sequence 73, Application US/08464000
; Patent No. 6335020
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce
; APPLICANT: Klapper, David G.
; APPLICANT: Rafnar, Thorunn
; APPLICANT: Kuo, Mei-chang
; TITLE OF INVENTION: Allergenic Peptides from Ragweed Pollen
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,000
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandagouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-464-000-73

Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 130
Best Local Similarity: 21.66% Mismatches: 93
Query Match: 4.70% Indels: 93
DB: 4 Gaps: 16

US-10-010-050a-2 (1-346) x US-08-464-000-73 (1-1349)
Qy      65 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 77
Db      526 TGTCACAGAGAGCATGATTAAAGTCCAAAGATGCTCAACATTTTAAACAAACAAAGTAT 585
Qy      78 AspAspIleGluValPheArgLeuGlnAlaProValTrrGluPheIysTyrGlyAspLeu 97
Db      586 GGTGATGCTTAATGATGTCGTGATGTTCAAAATATG----- 624
Qy      98 LeuGlyHisIleuIysIleMetHisAspAlaIleGlyPheArgSerThrlleuThrIlyIys 117
Db      625 ATGCACCATGCTGCTCAGTAGAGCTTCCAGATGAGCTGCTCCATATCACCCCTGGCAGC 684
Qy      118 AsnTyrThrMetGluTrrPylsGluPheGlnIleuGlyAsnGlySthrPheProHisIleu 137
Db      685 TCACACGTG-----ACCGTTTCCAACTGCAGAAATTC----- 714
Qy      138 ArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
Db      715 -----ACCCAAACACCAATTTGATTATTGCTGGG 744
Qy      158 IleAspAspValHisTrrPylsGluAsnGlyThrlleuValGlnValAlaThrIleSergly 177
Db      745 GCTGATGACACCCATTATCAAGATTAAGGAGATGTA-----GCAACGGTAGCATTC 795
Qy      178 AsnMetPheAsnGlnMetAlaLysTrrValIleGlnAspAsnGluThrGlyIleTyr--- 196
Db      796 AACATGTTCAACCATCACTGACGTTGACCAAGATGCTGATGATTTGGGTTTCCAA 855
Qy      197 -----TyrGluThrTrrAsnValIleAlaSerProGluIlySglValGlu 211
Db      856 GTCGTTAACAAACAACACTACGACAGATGGGAAACGTACCGTGGTAGCTGGCCCA 915
Qy      212 Thr-----TrrPheAspSerTyrAspCysSerIysPheValIleu 224

```

```

b 916 ACTATACCTCAGCCAGGAAAGATCTCTGCCCCCGCATGATCATCAG----- 966
y 225 ArgThrheAsnyslsuAlaGluPheGly-----AlaGluPheAsnileGlu 241
b 967 -----AAAAATGCTTTCAGGAGACTGCTGTTCTGCAACGAGAGTGCATCTGGAAC 1420
y 242 ThrAsnTyThrArgIlePheLeuTyrSerGlyGluProThyTyrLeuGlyAsnGluThr 261
b 1021 TGGAGAACAGATAGAGACTCTCTGAAATGCT----- 1053
y 262 SerValPheGlyProThyGly-----AsnysThrLeuGlyLeuAlaIleLysArgPheTyr 280
b 1054 GCTATTTTCTCCCATCCGGCTGATCCAGTGTACACCCCTGAGCAAAAAGCAGGATG 1.13
y 281 TyrProPheLysProHisLeuProThyGluPheLeuSerLeuLeuGluIlePhe 300
b 1114 ATTCAGCTGACCA-----GAGAGAGCCGTTCTTAAGACTCAGTACAGTACGCT 1.61
y 301 AspAlaValIleValHisLys-----GlnPheTyrLeuPhe 312
b 1162 GGTGACTCTCATGCCATCAGAGACCTTGCTAAGCAGCCTGGCCATTCCTAAGCTTT 1221
y 313 TyrAsn-----PheGluTyrTrpPheLeuProMetLysPheProPheIleLysIle 329
b 1222 TATATATATCAATTAATCTTATTATTATTAT-----TTGATATAT 1263
y 330 ThrTyrGluGluIleProLeuProIleArgAsnysThrLeu 343
b 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293

```

RESULT 10

US-08-480-528A-11

Sequence 11, Application US/08480528A

Patent No. 5652118

GENERAL INFORMATION:

APPLICANT: OPPERMAN, HERMANN

APPLICANT: OZAKAYAK, ENGIN

APPLICANT: KUBERASAMPATH, THANAGAVEL

APPLICANT: RUBER, DAVID C.

APPLICANT: PANG, ROY H. L.

APPLICANT: COHEN, CHARLES M.

TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES

STREET: 45 SOUTH STREET

CITY: HOPKINTON

STATE: MA

COUNTRY: USA

ZIP: 01748

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,528A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: FENTON Esq., GILLIAN M.

REGISTRATION NUMBER: 36,508

REFERENCE/DOCKET NUMBER: CRP-076FW

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7560

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 6418 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..6361

OTHER INFORMATION: /note= "HOP-2 genomic sequence"

FEATURE:

NAME/KEY: exon

LOCATION: 1..837

OTHER INFORMATION: /note= "EXON ONE"

FEATURE:

NAME/KEY: misc_feature

LOCATION: 884..885

OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN

OTHER INFORMATION: POSITIONS 884 AND 885 IN THIS SEQUENCE"

FEATURE:

NAME/KEY: exon

LOCATION: 1088..1277

OTHER INFORMATION: /note= "EXON TWO"

FEATURE:

NAME/KEY: exon

LOCATION: 1350..1814

OTHER INFORMATION: /note= "EXON THREE"

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1834..1835

OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN

OTHER INFORMATION: POSITIONS 1834 AND 1835 IN THIS SEQUENCE"

FEATURE:

NAME/KEY: exon

LOCATION: 1883..2077

OTHER INFORMATION: /note= "EXON FOUR"

FEATURE:

NAME/KEY: exon

LOCATION: 2902..2981

OTHER INFORMATION: /note= "EXON FIVE"

FEATURE:

NAME/KEY: exon

LOCATION: 3507..3617

OTHER INFORMATION: /note= "EXON SIX"

FEATURE:

NAME/KEY: exon

LOCATION: 6116..6361

OTHER INFORMATION: /note= "EXON SEVEN"

US-08-480-528A-11

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Pred. No.	Score	Percent Similarity	Best Local Similarity	Query Match	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
12.3	90.50	29.97%	4.70*	1	6418	67	28	97	126	14	

US-10-010-050A-2 (1-346) x US-08-480-528A-11 (1-6418)

13 setTPCystrTPAlaLeuAlaLeuLeuTrpLeuAlaValaValProGly----- 28

532 TCTGTGGGCTGCGCGCTATGCG-----CGCTGGGGGGGGGGGGGGCGCGCGAGACCCCGC 588

29 -----TTPSerArgValSer---GlyTLeProSerArgArgHis 40

589 CCGGCTGTCCCGACGACGCTGTGGCGCGCGAGCGCGGAGCGGAGCGGAGATCC 648

41 TTP-----ProValProTyrLysArgPheAspPheArgPro 52

649 TGGCGGTGCTCGGGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693

53 LysProAspProTyrCys-GlnAlaLysTyrThrPheCysProThnGlySerProIlePr 72

694 CCGCGCGCTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 747

72 C-ValMetGluGlyAspAspAsp----- 79

Db 748 AGCCATGCGCGGCGAGCAGCAGGAGGAGGCGCGCGCGGAGCGCGCTTGGGCGCGG 807
Qy 79 ----- 79
Db 808 CCGAGCTGTGATGAGCTTGTGTTACATGAGTGAGTGCGGCGCGCGGAGCGCTTGC 667
Qy 80 --11leGluValPheArgLeuGlnAlaProValTrrpGluPheArgTyrGlyAspLeuLeuG 59
Db 868 GAGTAACTGCTGCGAGCTGCGAGGCTCTTGTGGCTCTACACCGCGGAGCAGAGCTTG 527
Qy 99 1yHisLeuArg1LeuMetHisAspAlaTleGlyPheArgSerThrLeuThrGlyLysAsnT 119
Db 928 AACAAAGTTTGACATAAATGAAGAGCGCGCGCGCGCGCTTCTGCTCCAGC- 586
Qy 113 YrHmetGluTrrpTyrGluLeuPheGlnLeuGlyAsnGlySerThrPheProH1eLeuArgP 139
Db 987 -----CTTGAAGTGTGGTGAGTGC- TGG-----GGGCGGTGGCTCAACAGCTCTGCCCC 1034
Qy 139 roGluMetAspAlaProPhe-TrrpCysAsnGlnGlyAlaAla-CysPhePheGlu----- 156
Db 1035 CTCCAGAGCGCGAGCGCTTGTGAGTGCGAGCGCGCGCTTGTCTTGTCTTGTAGTGAGCG 094
Qy 157 -----GlyTleAspAspValHisTrrpLysGlu----- 65
Db 1095 AGACCGTGCCTGGGCGACCGACCGCGCGCGCGCGCGCTTGTGAGCTGAGCGCA 154
Qy 166 -----AsnGlyThrLeuValGlnValAla----- 73
Db 1155 GATCCCGGCTGGGAGGCGGCTGACAGCTGCGAGTTCGGATTTACAGGTGCCAGCAT 1214
Qy 174 -----ThrIleSerGlyAsnMetPheAsnGlnMet----- 183
Db 1215 CCAGCTGTCAACAGAGAGCGCTCCACGTCAGATGTTCCAGGTGTTCAGAGAGCATCCAA 1274
Qy 184 -----AlaLysTrrpVa 187
Db 1275 CAGGTGCTTCCCTTGGCGCGGCGCGCGCGCGCGCGCGCGCGCTTGTATGATGTC 1334
Qy 187 1lyGlnAspAsnGlnThrGlyLeuTyrTrrpGluThrTrrpAsnValLysAlaSerProG 207
Db 1335 AAGCAGCGCGCGAGG-----GAGTGTGGTGGTGAAGAGAGCGCTCA 1379
Qy 207 uLysGlyAlaGluThrTrrpPheAspSerTrrpAspCysSerLys 221
Db 1380 AGATGGGAAGATGCTTGGCGGAGCGCGCTTGCAGCTGTGGAG 1422

RESULT 11
US-08-479-666-11
; Sequence 11, Application US/08479666
; Patent No. 5652337
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H. L.
; APPLICANT: COHEN, CHARLES M.
; TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,666

FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FENTON Esq., GILLIAN M.
REGISTRATION NUMBER: 36,508
REFERENCE/DOCKET NUMBER: CRP-076DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 435-9001
TELEFAX: (508) 435-6951
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 6418 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..6361
OTHER INFORMATION: /note= "top-2 genomic sequence"
FEATURE:
NAME/KEY: exon
LOCATION: 1..837
OTHER INFORMATION: /note= "EXON ONE"
FEATURE:
NAME/KEY: misc feature
LOCATION: 884..885
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
OTHER INFORMATION: POSITIONS 884 AND 885 IN THIS SEQUENCE"
FEATURE:
NAME/KEY: exon
LOCATION: 1088..1277
OTHER INFORMATION: /note= "EXON TWO"
FEATURE:
NAME/KEY: exon
LOCATION: 1350..1814
OTHER INFORMATION: /note= "EXON THREE"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1834..1835
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
OTHER INFORMATION: POSITIONS 1834 AND 1835 IN THIS SEQUENCE"
FEATURE:
NAME/KEY: exon
LOCATION: 1883..2077
OTHER INFORMATION: /note= "EXON FOUR"
FEATURE:
NAME/KEY: exon
LOCATION: 2982..2981
OTHER INFORMATION: /note= "EXON FIVE"
FEATURE:
NAME/KEY: exon
LOCATION: 3507..3617
OTHER INFORMATION: /note= "EXON SIX"
FEATURE:
NAME/KEY: exon
LOCATION: 6116..6361
OTHER INFORMATION: /note= "EXON SEVEN"
US-08-479-666-11

Alignment Scores:
Pred. No.: 12.3 Length: 6418
Score: 90.50 Matches: 67
Percent Similarity: 29.97% Conservative: 28
Best Local Similarity: 21.14% Mismatches: 97
Query Match: 4.70% Indels: 126
DB: 1 Gaps: 14

US-10-010-050A-2 (1-346) x US-08-479-666-11 (1-6418)
Qy 13 serTrrpCysTrrpAlaLeuAlaLeuLeuTrrpLeuAlaValAlaProGly----- 28
Db 532 TCTGTGGCTGTGGCGCTGATGCG---CGTGGGCGGCGGCGCGCGCGCGCTGCGAGCGCGCG 588

29 -----TipSerArgValSer---GlyIleProSerArgHis 40
 589 CCGGCTGCCCCAGCAGCTCTGGGCGCGCCGAGCGCCGAGCGAGATCC 648
 41 TTP-----ProValProTyrIlysArgPheAspPheArgPro 52
 649 TGGCGGTGCTGGGCTGCTGGCGCGCGCCCGCCCGCC-----GCCGCCAC 693
 53 LysProAspProTyrCys-GlnAlaIysTyrPheCysProThrGlySerProIlePr 72
 694 CCGCGCTCCCGGCTGCGCGCGCGCGCTCTT-----CATGCTGACCTGTAAC 747
 72 O-ValMetGluGlyAspAspAsp-----79
 748 AGCGCATGCGCGCGAGCAGCAGCAGCAGCGCGCGCCGCGAGCGCGCTGGCGCG 807
 79 -----79
 808 CCGACCTGTCATAGACTTCTTAACTAGTGTGAGTGGCGCGCGCGCGGAGACCTCG 867
 80 --IleGluValPheArgLeuGlnAlaProValTyrGluPheIysTyrGlyAspLeuEng 93
 868 GAGTAACTGGCTGAGCTGAGCGCGCTCTTCTGCTCTACACCCCGGAGCCAGCTCG 927
 99 LysIleuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyIysAsnT 119
 928 AACAAACGTTGCACTAATGATGAGCGCGCGCCAGCCAGCCCTCCCTGGCTCCGCTCCAC- 936
 119 YrThrMetGluTyrPyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgP 139
 987 -----CTTGAGTGGTGGGGTGGC-TGG-----GGCGGTGGCTCACACAGCTGTGCCCC 1334
 139 roGluMetAspAlaProPhe-TripCysAsnGlnGlyAlaIle-CysPhePheGlu----- 136
 1035 CTCGACAGAGCCGAGCATTTCTGAGTGCAGCGCCAGCGGTGCTTCTTCTTCTAGTGAAGG 1094
 157 -----GlyIleAspAspValHisTyrIysGlu----- 165
 1095 AGACCGTGGCTGGGCGCACAGAGCGCCCATTTGAGAGATTCGCGCTTGACCTGACCCA 1154
 166 -----AsnGlyThrLeuValGlnValAla----- 173
 1155 GATCCCGGCTGGAGCGGCTCACAGCTGGCGAGTTCCGGATTACAGAGTCCCAAGCAT 1214
 174 -----ThrIleSerGlyAsnMetPheAsnGlnMet----- 183
 1215 CCACCTGCTCAACAGAGACCTTCACGTCGACATGTTCCAGTGTGTCAGAGACAGTCCA 1274
 184 -----AlaIysTyrPva 187
 1275 CAGGTGCTTCCCTTGGCGCGGTCGCCAACCCTAACCCCTCAACAGTCTCATGTGC 1334
 187 LysGlnAspAsnGlnThrGlyIleTyrTyrGluThrPanValIysAlaSerProI 207
 1335 AAGGCGAGCCAGAGG-----GAGTGTGTGGTGTAAAGAGAGCTCTCA 1379
 207 uLysGlyAlaGluThrTyrPheAspSerTyrAspCysSerIys 221
 1380 AGATGGGAAGATGCTTGGCCGAGGCGCTGCACCTGGGAAG 1422

COUNTRY: USA
 ZIP: 01748
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/10520
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/667,274
 FILING DATE: 11-MAR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/752,764
 FILING DATE: 30-AUG-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/753,059
 FILING DATE: 30-AUG-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/752,857
 FILING DATE: 30-AUG-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/923,780
 FILING DATE: 31-JUL-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/922,813
 FILING DATE: 31-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: PITCHER ESQ, EDMUND R
 REGISTRATION NUMBER: 27,829
 REFERENCE/DOCKET NUMBER: CRP-076PC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508)435-9001
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6418 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: 1..6361
 OTHER INFORMATION: /note= "HOP-2 genomic sequence"
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1..837
 OTHER INFORMATION: /note= "EXON ONE"
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: 884..885
 OTHER INFORMATION: /note= "A Gap Occurs Between
 OTHER INFORMATION: Positions 884 and 885 in this Sequence"
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1088..1277
 OTHER INFORMATION: /note= "EXON TWO"
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1350..1814
 OTHER INFORMATION: /note= "EXON THREE"
 NAME/KEY: misc feature
 LOCATION: 1834..1835
 OTHER INFORMATION: /note= "A Gap Occurs Between
 OTHER INFORMATION: Positions 1834 and 1835 in this Sequence"
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1883..2077
 OTHER INFORMATION: /note= "EXON FOUR"
 FEATURE:

RESULT 12
 PCT-US93-10520-11
 Sequence 11, Application PC/TUS9310520
 GENERAL INFORMATION:
 APPLICANT:
 TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CREATIVE BIOMOLECULES, INC.
 STREET: 45 SOUTH STREET
 CITY: HOPKINTON
 STATE: MA


```

APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1328
US-08-290-448A-58
Alignment Scores:
Pred. No.: 1.23 Length: 1328
Score: 89.50 Matches: 68
Percent Similarity: 29.45% Conservative: 28
Best Local Similarity: 20.86% Mismatches: 113
Query Match: 4.64% Indels: 117
DB: 1 Gaps: 15
US-10-010-050a-2 (1-346) x US-08-290-448A-58 (1-1328)
QY 65 CysPro-----ThGlySerProIleProValMetGluGlyAsp 77
DB 505 TGTCCAGAGGCGATTAAGTCCACAGATGTCACCAATTAAAGACAAAGAGGAT 364
QY 78 AsnAspIleGluValPheArgGluGlnAlaProValTrpGluPheLysTyGlyAspLeu 37
DB 565 GGTATGCTATAATGTTGCTGGTAGTTCACAAATATCG----- 503
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
DB 604 ATGCACCATGCTGCTCAGTAAGGCTCCGATGGCTGCATATCACCCCTGGCAGC 663
QY 118 AsnTyThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 137
DB 664 TCACACGCGT-----ACGGTTCACCACTGCMAATTC----- 693
QY 138 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
DB 694 -----ACCCAAAGCAAAATTGATTATGCTGGG 723

```

```

QY 158 IleAspAspValHisTrpTyrGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
DB 724 GCTGATGACACCCATTATCAAGATTAAGGATGCTA-----GCAACGGTATGACTTC 774
QY 178 AsnMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTy 196
DB 775 AACATGTTCCACCGATCAAGTTGACCCAAAGATGCTAGATGTAGATTGGGTTTCCAA 834
QY 197 -----TyrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGlu 211
DB 835 GTGCTTAACAACACTACGACAGATGGGAAACGTACGCCATCGGTAGTGCGCCCA 894
QY 212 Thr-----TrpPheAspSerTyrAspCysSerLys----- 221
DB 895 ACTATACACGACCAAGGACACAGATTTCTGGCCCCGATGATATCATCAAGAAAAATGTC 954
QY 222 -----PheValLeuArgThrPheAsn 228
DB 955 TTAGCGAGGACTGTACTGCGAACGACAGATCGATGCTGGAAGTGGAAACAGATAA 1014
QY 229 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 248
DB 1015 GACTTGCTTGAAAATGAGTCT-----ATTTT 1041
QY 249 LeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 268
DB 1042 CTCACATCGGGCTCGATCCAGTCACTGCTAACCCCTGAGCAA----- 1080
QY 269 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuPro 288
DB 1081 -----AAAGCAGGAGATGTTCCAGCTGAACA----- 1107
QY 289 ThrLysGluPheLeuLeuSerLeuGlnIlePheAspAlaValIleValHisLys 307
DB 1108 --GGAGAACCGCTTAAAGACTCACTAGTACGTGCGTACTCTCCATGCATCAAGGA 1164
QY 308 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 317
DB 1165 GCACCTTGCTAAGACACTGGCCCAATTCCTAAGCTTTATATATATCAATAACTATATT 1224
QY 318 TrpPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 337
DB 1225 TATTATTAT-----TTGATATTATTATATGAA-----CCA 1254
QY 338 IleArgAsnLysThrLeu 343
DB 1255 TTACGTTCAAGTACTCTA 1272
RESULT 15
US-08-290-448A-58
Sequence 58: Application US/08290448A
Patent No. 5698204
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A

```

FILING DATE: August 15, 1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/529,951
 FILING DATE: May 29, 1990
 APPLICATION NUMBER: US 07/325,365
 FILING DATE: March 17, 1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Amy E. Mandragouras
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: IM1-018CN
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)227-5941
 INFORMATION FOR SEQ ID NO: 58:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1328 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1328
 US-08-290-448A-58

Alignment Scores:
 Pred. No.: 1.23 Length: 1328
 Score: 89.50 Matches: 68
 Percent Similarity: 29.45% Conservative: 28
 Best Local Similarity: 20.86% Mismatches: 113
 Query Match: 4.64% Indels: 117
 DB: 1 Gaps: 15

US-10-010-050A-2 (1-346) x US-08-290-448A-58 (1-1328)

QY 65 CysPro-----ThrglySerProIleProValMetGluGlyAsp 77
 |||||
 DB 505 TGTCCAGGAGGAGGATTAAGTCCACAGATGCTCCACCAATTTTAAGACAACAAAGATGAT 364
 QY 78 AspAspIleGluValPheArgGluLeuAlaIleProValTrrpGluPheLysTyrGlyAspLeu 97
 |||||
 DB 565 GGTGATGCTATAAATGTTGCTGTAAGTTCACAAATATGG----- 603
 QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
 |||||
 DB 604 ATCCAGCCATGTGCTCCCTCAGTAAGGCTTCAGATGGCTGCTGATATCAACCTCGGACAGC 663
 QY 118 AsnTyrThrMetGluTrrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 137
 |||||
 DB 664 TCACACGGT-----ACCGTTCCAACTGCATAATTC----- 693
 QY 138 ArgProGluMetAspAlaProPheTrrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
 |||||
 DB 694 -----ACCCAAACCAATTTGATTAATGCTCGGG 723
 QY 158 IleAspAspValHisTrrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
 |||||
 DB 724 GCTGATGACACCCATTAATACAGATAAGCATGCTA-----GCAACGCTAGCATTC 774
 QY 178 AsnMetPheAsnGlnMetAlaLysTrrpValLysGlnAspAsnGluThrGlyTleTyr--- 196
 |||||
 DB 775 AACATGTTACGATCAGCTTGACCAAGATGCTAGATGTAATTTGGGTTTTCCAA 834
 QY 197 -----TyrGluThrTrrpAsnValLysAlaSerProGluLysGlyAlaGlu 211
 |||||
 DB 835 GTGCTTAACACAACTACACAGATGGGAGACGTACGCTGCTGGGCCCA 894
 QY 212 Thr-----TrrpPheAspSerTyrAspCysSerLys----- 221
 |||||
 DB 895 ACTATATCTAGGCCAAGGAGACAGATTTCTGCCCCGATGATATCATCAAGGAAATGTC 954
 QY 222 -----PheValLeuArgThrPheAsn 228
 |||||

DB 955 TTAGCGAGAGCTGCTACTGCGACAGCGACAGTCAGTGTGGAACTGGAGACAGATAAA 1014
 QY 229 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 248
 |||||
 DB 1015 GACTTGCTGGAATAATGGTGCT-----ATTTT 1041
 QY 249 LeuTyrSerGlyGluProThrTyrThrLeuGlyAsnGluThrSerValPheGlyProThrGly 268
 |||||
 DB 1042 CTCCCATCCGGGTCTGATCCAGTGTCAACCCCTGAGCAA----- 1080
 QY 269 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTrrpPheLysProHisLeuPro 288
 |||||
 DB 1081 -----AAAGCAGGAGATTCACGCTGAACCA----- 1107
 QY 289 ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLys--- 307
 |||||
 DB 1108 ---GGAGAAAGCGGTCTTAAGACTCACTAGTAGTGTGGTGTACTCTCATGCGCATCAAGGA 1164
 QY 308 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 317
 |||||
 DB 1165 GCACCTTGCTAAGCAACCTGCGCAATTCCTAAGCTTTTATTAATATCATTAATCTATTT 1224
 QY 318 TrrpPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 337
 |||||
 DB 1225 TATTTATTT-----TTTGATATTTTATATGAA-----CCA 1254

Search completed: April 25, 2004, 05:50:25
 Job time: 148.853 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd

```

JM protein - nucleic search, using frame_plus_p2n model
Run on:      April 24, 2004, 22:59:01 ; Search time 638.658 Seconds
              (without adjustments)

```

```

Title: US-10-010-050A-2
Perfect score: 1927
Sequence: 1 MRRGAGAGRRASWCWALAL.....IKITYEETPLPIRNTLSGL 346

```

Scoring table:		BLOSUM62
Xgapop	10.0	Xgapext 0.5
Ygapop	10.0	Ygapext 0.5
Fgapop	6.0	Fgapext 7.0
Delop	6.0	Delext 7.0

```
Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

Command line parameters:
-MODEL=framem+ p2n.model -DEV=xlh
-O=/cgm2.1/USPFO.spool/US100100050/runat.22042004.113203.27539/app_query.fasta_1.1372
-DB=Genesec.291a04/-GEMT-fastlap -SUFFIX=ring -MINMATCH=0.1 -LOOCTL=0
-LOOEXCT=0 -UNITS=bits -STAR=1 -END=-1 -MATTRX=10sum62 -GRNS=humana0.cdi
-LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMF=apo -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US100100050/TCGN_1.1.819 @runat.22042004.113203.27539 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-NO TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database :
N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1927	100.0	1486	2	AAK02855	Aax02855 Human zsl
2	1916	99.4	1751	2	AAK24826	Aax24826 Human sec
3	1913	99.3	2120	2	AAK94829	Aak94829 Human ful
4	1631	84.6	1038	2	AAK02866	Aax02866 Human ceg
5	1065	55.3	697	4	AAK93424	Aak93424 Human cdn
6	1065	55.3	697	4	AAK92158	Aak92158 Human cdn
7	886	46.0	494	4	AAK25096	Aak25096 Probe #15
8	886	46.0	494	4	ABA70786	Aba70786 Human foe

9	886	46.0	494	4	AA150961
10	886	46.0	494	4	ABR37272
11	886	46.0	494	4	AAK44992
12	886	46.0	494	4	AAK13040
13	886	46.0	494	4	ABSA4658
14	886	46.0	494	6	ABSI1937
15	727	37.7	506	2	AAV88347
16	684.5	35.5	474	4	AA115879
17	684.5	35.5	474	4	ABAS8185
18	684.5	35.5	474	4	AA137794
19	684.5	35.5	474	4	ABR27266
20	684.5	35.5	474	4	AAK31927
21	684.5	35.5	474	4	AAK06266
22	684.5	35.5	474	4	ABSI1617
23	684.5	35.5	474	6	ABSO6689
24	472	24.5	497	6	ABO55979
25	359.5	18.7	351	6	ABL79601
26	241	12.5	473	6	ABE66053
27	241	12.5	473	6	ABE68165
28	180	9.3	630	6	ABQ40653
29	180	9.3	631	6	ABQ40652
30	180	9.3	631	6	ABQ13702
31	180	9.3	631	6	ABQ13703
32	124	6.4	630	6	ABO40651
33	124	6.4	630	6	ABO40650
34	124	6.4	631	6	ABQ13704
35	124	6.4	631	6	ABQ13705
36	121	6.3	60	6	ABN37951
37	112	5.8	2120	4	AA160668
38	112	5.8	2175	4	AAH16737
39	103	5.3	1634	3	AAAC5994
40	99.5	5.2	4957	4	AAK85867
41	96	5.0	3580	6	AAAL6637
42	96	5.0	3580	6	ABR35630
43	95	5.0	4366	4	AAH57556
44	95.5	5.0	3567	5	AAAC91905
45	95.5	5.0	4858	5	AAAC91904
					AA150961 Probe #19
					ABR37272 Probe #19
					AAK44992 Human bon
					AAK13040 Human bra
					ABSA4658 Human liv
					ABSI1937 Human gen
					AAV88347 EST clone
					AA15879 Probe #58
					ABO55979 Human foe
					AA137794 Probe #54
					ABR27266 Probe #58
					AAK31927 Human bon
					AAK06266 Human bra
					ABSI1617 Human liv
					ABSO6689 Human gen
					ABQ55979 Human ova
					ABL79601 Human ova
					AB160053 Lung cano
					ABE68165 Kidney can
					ABQ40653 Oligonuc
					ABQ40652 Oligonuc
					ABQ13702 Oligonuc
					ABQ40651 Oligonuc
					ABQ40650 Oligonuc
					ABQ13704 Oligonuc
					ABQ13705 Oligonuc
					ABN37951 Human sp
					AA160668 Human po
					AAH16737 Human CD
					AAAC5994 Human se
					AAK85867 Human im
					AAAL6637 Human se
					ABR35630 cDNA seq
					AAH57556 Human br
					AAAC91905 Murine A
					AAAC91904 Murine A

ALIGNMENTS

RESULT 1
AAX02855
ID AAX02855 standard; DNA; 1486 BP.

AC AAX02855

DT 14-MAY-1999 (first entry)

Human zsig46 DNA.

secreted protein; zsig46; human; chromosome 13; thyroid disease; hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer; Hirschsprung's disease; neuronal ceroid-lipofusiosis; Wilson disease; Reiger syndrome; immunassay; detection; anti-idiotypic antibody; therapy; diagnostic; ss.

OS Homo sapiens.

Key	Location/Qualifiers
FH	7 1007

PN W09905275-A1

PD 04-FEB-1999

PF 24-JUL-1998; 98WO-US015431

PR 24-JUL-1997; 97US-0053613P

PA (ZYMO) ZYMOGENETICS INC.

XX Sheppard PO, Gilbertson DG;
 PI
 XX
 DR MPI; 1999-142930/12.
 P-PSDB; AAM92967.
 XX
 PT New secreted polypeptide, zsig46, and its fragments, related fusion
 PT proteins - used for diagnosis and treatment of thyroid disorders or
 PT diseases involving genes on chromosome 13.
 XX
 PS Claim 27, Page 88-90; 101pp; English.
 XX
 CC This invention describes the isolation of a novel human secreted protein,
 CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in
 CC the thyroid. This product can be used to study secretion of proteins from
 CC cells and also to treat or prevent deficient expression of zsig46, which
 CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
 CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
 CC involve genes in the same region of chromosome 13 (e.g. Hirschprung's
 CC disease, neuronal ceroid-lipofusiosis, Wilson disease and Reiger
 CC syndrome). Antibodies and other binding proteins, are used as immunoassay
 CC reagents to detect zsig46 or cells expressing it, e.g. for assessing
 CC thyroid function to produce anti-idiotypic antibodies, for affinity
 CC purification of zsig46, to screen expression libraries, to neutralise
 CC zsig46 activity, and to deliver toxins, radioisotopes etc. for
 CC therapeutic or diagnostic purposes. Agonists of the product can be used
 CC to promote growth, differentiation and proliferation of specific cell
 CC types, e.g. for treating (extra)thyroid diseases or as additive to cell
 CC cultures
 CC
 XX
 SO Sequence 1486 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;
 SO
 Alignment Scores:
 Pred. No.: 4.89e-194 Length: 1486
 Score: 1927.00 Matches: 346
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-10-010-050A-2 (1-346) x AAK02855 (1-1486)
 QY 1 MetProValProTyrTyrLeuArgPheAspPheArgProLysProAlaProTyrCysGluAla 20
 Db 47 ATGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 106
 QY 21 LeuTrpLeuAlaValAlaProGlyTyrPheArgValSerGlyIleProSerArgArgHis 40
 Db 107 CTTTGCGCTCGGGGCTGCTCGGGGCTGCTCGGGGCTGCTCGGGGCTGCTCGGGGCTGCTCGGGG 166
 QY 41 TrpProValProTyrTyrLeuArgPheAspPheArgProLysProAlaProTyrCysGluAla 60
 Db 167 TGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 226
 QY 61 LysTyrTrpPheCysProThiGlySerProIleProValMetGluGlyAspAspAspIle 80
 Db 227 AAGTACTCTTGTCTCACTGCTCACTCTCACTCTCACTCTCACTCTCACTCTCACTCTCACTCT 286
 QY 81 GluValPheArgLeuGluAlaProValTyrGluPheLysTyrGlyAspLeuLeuGlyHis 100
 Db 287 GAAATTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 346
 QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerTrpLeuThiGlyLysAsnTyrThr 120
 Db 347 TTGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406
 QY 121 MetGluTyrPyrGluLeuPheGluLeuGlyAsnCysTrpPheProHisLeuArgProGlu 140
 Db 407 ATGGATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 466
 QY 141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGlyIleAspAsp 160
 Db 467 ATGGATGCCCTTTTGGTGTATCAAGCGCTGCTGCTTTTGGAGGAATGATGAT 526

QY 161 ValHisTrpLysGluAsnGlyThrLeuValGluValAlaThrIleSerGlyAsnMetPhe 180
 Db 527 GTTCACTGGAAGAAATGGACATTAGTCAAGTAGCACTATATCAGAAACATGTTTC 586
 QY 181 AsnGlnMetAlaLysTrpValLysGluAsnGluThrGlyIleTyrTyrGluThrTrp 200
 Db 587 AACCAATGGCAAGTGGTGAACACGACAAATGAAGAAAGAAATTTATTAAGACATGG 646
 QY 201 AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSer 220
 Db 647 AATGTAAAGCCAGCCAGAAAGGGGCGACAGACATGTTGATTCCTCAGCATGTTCC 706
 QY 221 LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240
 Db 707 AATTTGCTTAAAGACCTTTTAAAGTTGCTGAATTTGGACAGAGTTCAAGAACATA 766
 QY 241 GluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGlu 260
 Db 767 GAAACCAACTATACAGAAATTTCTTTACAGTGGAGAACCTACTTATCTGGGAAATGAA 826
 QY 261 ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
 Db 827 AACTCTGTTTGGCGCAACAGAAACAGACTCTTGTGTTAGCCATAAAGATTATAT 886
 QY 281 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuLeuGlnIlePhe 300
 Db 887 TACCCCTTCAACCACTATGGCACTAAAGAAATTTCTGTAGTCTTGCAAAATTTT 946
 QY 301 AspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu 320
 Db 947 GATGACGATATGTGACAAACAGTCTATTTGTTTAAATTTTAAATTTTAAATTTTAA 1006
 QY 321 ProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsn 340
 Db 1007 CCTATTAATTCCTTTTATTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 1066
 QY 341 LysThrLeuSerGlyLeu 346
 Db 1067 AAAACACTCTCTGTTTA 1084
 Db
 RESULT 2
 AAZ24826
 ID AAZ24826 standard; DNA; 1751 BP.
 XX
 AC AAZ24826;
 XX
 DT 02-DEC-1999 (first entry)
 DE
 DE Human secreted protein gene 16 clone HMZAD77.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN W09947540-A1.
 PD 23-SEP-1999.
 XX
 EF 18-MAR-1999; 99WO-US005804.
 XX
 PR 19-MAR-1998; 98US-0078563P.
 PR 19-MAR-1998; 98US-0078566P.
 PR 19-MAR-1998; 98US-0078573P.
 PR 19-MAR-1998; 98US-0078574P.
 PR 19-MAR-1998; 98US-0078576P.

(HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI: 2001-524255/58.
DR P-PSDB: AAM93870.

XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.

PS Claim 8; SEQ ID NO 3977; 1380bp + Sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5' and 3' ends of the cDNA molecules are useful
CC for determining primers for synthesizing the full length cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesized by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from BPO

XX Sequence 2120 BP; 590 A; 409 C; 460 G; 661 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,53e-192 Length: 2120
Score: 1913.00 Matches: 344
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 2
Query Match: 99.27% Indels: 0
DB: 4 Gaps: 0

US-10-010-050a-2 (1-346) x AAK94829 (1-2120)

QY 1 MetArgArgGlyAlaGlyAlaAlaArgAlaSerThrProCysTrpAlaLeuAlaLeu 20
DB 52 ATGGCGCGGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 111
QY 21 LeuThrPheAlaValAlaProGlyTyrSerArgValSerGlyLeuProSerArgArgHis 40
DB 112 CTTGGCTCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 171
QY 41 TrpProValProTyrLysArgPheAspPheArgProLysProAspProTyrCysGlnAla 50
DB 172 TGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 231
QY 61 LysTyrThrPheCysProThrGlySerProLysProLysMetGlyLysAspAsp1le 80
DB 232 AAGTAATACCTTCTGCTCCAACTGGCTCACCATTACCACTTATGAGAGGTGATGACACT 291
QY 81 GluValPheArgLeuGlnAlaProValTyrGlnPheLysTyrGlyAspLeuGlyHis 100
DB 292 GAACTTTTCGATTCACAGCCCGCACTATGAGGAATTTAATGAGAACCTCTCGGACAC 351
QY 101 LeuLysIleMetHisAspAlaAlaGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
DB 352 TTGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411
QY 121 MetGluTyrTyrGluLeuPheGlnLeuGlyLysCysThrPheProHisLeuArgProGlu 140
DB 412 ATGGATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 471
QY 141 MetAspAlaProPheThrCysAsnGlnGlyAlaAlaCysPhePheGlnGlyTLeuAsp 160
DB 472 ATGGATGGCCCTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 531
QY 161 ValHisTyrPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 180
DB 532 GTTCACTGGAG 591

QY 181 AsnGlnMetAlaLysTyrPheValLysGlnAspAsnGlnThrGlyTyrTyrGluThrTrp 200
DB 592 AACCAATATGCAAGTGGGTGAAACAGACATGAAACAGAAATTTATTAAGACATGG 651
QY 201 AsnValLysAlaSerProGlnLysGlyAlaGlnThrTrpPheAspSerTyrAspCysSer 220
DB 652 AATGTAAAGCCAGCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 711
QY 221 LysPheValLeuArgThrPheAsnLysLeuAlaGlnPheGlyAlaGlnPheLysAsn1le 240
DB 712 AATTTGCTTAAAGACCTTTTAAAGTGGCTGAATTTGAGCAGAGTTCAAGAACATA 771
QY 241 GluThrAsnTyrThrArgTyrPheLeuTyrSerGlyGlnProThrTyrLeuGlyAsnGlu 260
DB 772 GAAACCAACTATACAGAAATATTTCTTAAAGTGAAGACCTACTATCTGGAAATGAA 831
QY 261 ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
DB 832 ACATCTGTTTGGGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891
QY 281 TyrProPheLysProHisLeuProThrLysGlnPheLeuSerLeuLeuGln1lePhe 300
DB 892 TACCCCTTCAACACATTTGCCAATAAGAAATTTCTGTGAGTCTCTGCAGAAATTTT 951
QY 301 AspAlaValIleValHisLysGlnPheTyrLeuPheTyrAspPheGlnTyrTrpPheLeu 320
DB 952 GATGACAGTATGATGACAAACAGATTCATTTGTTTATTAATTAATTAATTAATTA 1011
QY 321 PrometLysPheProPheLysLysLysLysLysLysLysLysLysLysLysLysLys 340
DB 1012 CCTATGAAATTCCTTTTATTAATAACATTAAGAAATCCCTTACTATCAGAAAC 1071
QY 341 LysThrLeuSerGlyLeu 346
DB 1072 AAAACACTCTCTGCTTAA 1089

RESULT 4

AAK02866
ID AAK02866 standard; DNA; 1038 BP.

XX AAK02866;

XX 14-MAY-1999 (first entry)

XX AC

XX DT

XX XX

XX DE

XX XX

XX OS

XX Homo sapiens.

XX PN

XX W09905275-A1.

XX PD

XX 04-FEB-1999.

XX PF

XX 24-JUL-1998; 98WO-US015431.

XX PR

XX 24-JUL-1997; 97US-0053613P.

XX PA

XX (ZYMO) ZYMOGENETICS INC.

XX PI

XX Shepard PO, Gilbertson DG;

XX DR

XX WPI: 1999-142930/12.

XX PT

XX New secreted polypeptide, zsig46, and its fragments, related fusion
PT proteins - used for diagnosis and treatment of thyroid disorders or
PT diseases involving genes on chromosome 13.

PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-483447/52.

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.

XX PS Claim 4; SEQ ID NO 19091; 639pp + Sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2,45e-84	Length:	494
Score:	886.00	Matches:	163
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.98%	Indels:	0
DB:	4	Gaps:	0

US-10-010-050A-2 (1-346) x ABA70786 (1-494)

```

QY 184 AAlaySTPValLyVgInaSPaSnGluThrGlyIleTyTYrGluThrTrpAsnVallys 203
Db 3 GCAAAAGTGGGTGAACAGACGACATGAAACAGAAATTTATGTAGACATGAAATGTAAA 62
QY 204 AAlaserProGluVgLyAlaGluThrTrpheaSPSerTYrAspCYseSerIysPheVal 223
Db 63 GCCACCCAGAAAGAGGGGCGAGACATGCTTGTATCTCTACGACTGTTCCAAATTTGTG 122
QY 224 LeuAtgThrPheaSnLyLeuAlaGluPheGlyAlaGluPheLyAsnIleGluThrAsn 243
Db 123 TTAAAGACCTTTAAACAAGTTGGCTGAATTTGAGAGAGGTTCAAGAACTAGAAACCAAC 182
QY 244 TyrThrArgIlePheLeuTySerGlyGluProThrTYrLeuGlyAsnGluThrSerVal 263
Db 183 TATACAGAAATATTTCTTTACAGTGAACCTACTTATCTGGGAAATGAAACATCTGTT 242
QY 264 PheGlyProThrGlyAsnLySerThreugLyLeuAlaIleLyAsnGlyPheTYrProhe 283
Db 243 TTGGGCCCAAGAGAAACAGACTCTGGTTTACCATTAAGATTTTATATACCCCTTC 302
QY 284 LysProHisLeuProThrysgIupheLeuLeuSerLeuLeuGluIlePheSpAlaVal 303
Db 303 AAACCACTTTGCCAACAATAAGAAATTTCTGTGAAGCTCTTCCAAATTTTATGCAAGG 362
QY 304 IleValHisLyGlnPheTyLeuPheTyAsnPhGluTyTrpPheLeuProheLy 323
Db 363 ATTTGTCACAAACAGTTCTATTTGTTTATATTTTGAATATTTGGTTTATACCTATGAAA 422
QY 324 PheProPheIleTyIleThrTYrGluGluIleProLeuProIleArgAsnLyThrLeu 343
Db 423 TTCCCTTTATTAATAATACATATGAAATCCCTTTACTATCATCAAGAAACAAACACATC 482
QY 344 SerGlyLeu 346
Db 483 TCTGTTTA 491

```

RESULT 9

AA150961
ID AA150961 standard; DNA; 494 BP.

XX AC AA150961;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #19647 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000663.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488997/53.

Human genome-derived single exon nucleic acid probes useful for analyzing
gene expression in human placenta.

Claim 25; SEQ ID NO 19647; 654pp; English.

The present invention relates to single exon nucleic acid probes (SENP).
The present sequence is one such probe. The probes are useful for
producing a microarray for predicting, measuring and displaying gene
expression in samples derived from human placenta. The probes are useful
for antenatal diagnosis of human genetic disorders

SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2,45e-84	Length:	494
Score:	886.00	Matches:	163
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.98%	Indels:	0
DB:	4	Gaps:	0

US-10-010-050A-2 (1-346) x AA150961 (1-494)

```

QY 184 AAlaySTPValLyVgInaSPaSnGluThrGlyIleTyTYrGluThrTrpAsnVallys 203
Db 3 GCAAAAGTGGGTGAACAGACGACATGAAACAGAAATTTATGTAGACATGAAATGTAAA 62
QY 204 AAlaserProGluVgLyAlaGluThrTrpheaSPSerTYrAspCYseSerIysPheVal 223
Db 63 GCCACCCAGAAAGAGGGGCGAGACATGCTTGTATCTCTACGACTGTTCCAAATTTGTG 122
QY 224 LeuAtgThrPheaSnLyLeuAlaGluPheGlyAlaGluPheLyAsnIleGluThrAsn 243
Db 123 TTAAAGACCTTTAAACAAGTTGGCTGAATTTGAGAGAGGTTCAAGAACTAGAAACCAAC 182
QY 123 TTAAGACCTTTAAACAAGTTGGCTGAATTTGAGAGAGGTTCAAGAACTAGAAACCAAC 182
Db 123 TTAAGACCTTTAAACAAGTTGGCTGAATTTGAGAGAGGTTCAAGAACTAGAAACCAAC 182
QY 244 TyrThrArgIlePheLeuTySerGlyGluProThrTYrLeuGlyAsnGluThrSerVal 263
Db 183 TATACAGAAATATTTCTTTACAGTGAACCTACTTATCTGGGAAATGAAACATCTGTT 242

```

Y 264 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLeuArgPheTyrTyrProPhe 283
 b 243 TTTGGGCCAACAAGAAACAAAGACTCTGTGGCTTAAAGATTTATTAACCCCTTC 302
 Y 284 LysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaVal 303
 b 303 AAACCAATTGGCCAACTAAAGAAATTTCTGTGAGTCTCTTGCAATTTTGATGCAGTG 362
 Y 304 IleValHisLysGlnPheTyrLeuPheTyrAsnPheGlnTyrTyrPheLeuProMetLys 323
 b 363 ATTGTGCACAAACAGTCTCTATTGTTTATATTTGAAATATGTTTAACTATGATAA 422
 Y 324 PheProPheIleLysIleThrTyrGluGlnIleProLeuProIleArgAsnLysThrLeu 343
 b 423 TTCCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACACTC 462
 Y 344 SerGlyLeu 346
 b 483 TCTGGTTTA 491
 RESULT 10
 ID ABA37272 standard; DNA; 494 BP.
 GC ABA37272;
 AT ABA37272 (first entry)
 23-JAN-2002 (first entry)
 Probe #15738 for gene expression analysis in human heart cell sample.
 Human; gene expression; heart; microarray; vascular system; probe;
 cardiovascular disease; hypertension; cardiac arrhythmia;
 congenital heart disease; ss.
 Homo sapiens.
 WO200157274-A2.
 09-AUG-2001.
 30-JAN-2001; 2001MO-US000666.
 04-FEB-2000; 2000US-0180312P.
 26-MAY-2000; 2000US-0207456P.
 30-JUN-2000; 2000US-00608408.
 03-AUG-2000; 2000US-00632366.
 21-SEP-2000; 2000US-0234687P.
 27-SEP-2000; 2000US-0236359P.
 04-OCT-2000; 2000GB-00024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-48899/53.
 Single exon nucleic acid probes for analyzing gene expression in human hearts.
 Claim 4; SEQ ID NO 15738; 530pp; English.
 The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,45e-84 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.98% Indels: 0
 DB: 4 Gaps: 0
 US-10-010-050a-2 (1-346) x ABA37272 (1-494)
 Y 184 AlAlyeTyrValLysGlnPheAsnGluThrGlyIleTyrTyrGluThrTyrAsnValLys 203
 b 3 GCAAGGGGTGAACAAGCAATGAAACAGCAATTATATGACATGAAATGTAATAA 62
 Y 204 AlAserProGluLysGlyAlaGluThrTyrPheAspSerTyrAspCysSerLysPheVal 223
 b 63 GCCAGCCCAAGAAAGGGGGGAGAGACATGGTGTGATCTTACGACGTCTCCAAATTTGTG 122
 Y 224 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn 243
 b 123 TTAAAGACCTTTAACAGAGTTGGCTGAATTTGAGCAGAGTTCAAGAACATGAAACCAAC 182
 Y 244 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 263
 b 183 TATCAAGAAATATTTCTTACAGTGAAGAACTTATCTGGAATGAAATGAAACATCTGTT 242
 Y 264 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPhe 283
 b 243 TTTGGGCCAACAAGAAACAAAGACTCTGTGTGATGACATTAAGATTTATTAACCCCTTC 302
 Y 284 LysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaVal 303
 b 303 AAACCAATTGGCCAACTAAAGAAATTTCTGTGAGTCTCTTGCAATTTTGATGCAGTG 362
 Y 304 IleValHisLysGlnPheTyrLeuPheTyrAsnPheGlnTyrTyrPheLeuProMetLys 323
 b 363 ATTGTGCACAAACAGTCTCTATTGTTTATATTTGAAATATGTTTAACTATGATAA 422
 Y 324 PheProPheIleLysIleThrTyrGluGlnIleProLeuProIleArgAsnLysThrLeu 343
 b 423 TTCCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACACTC 482
 Y 344 SerGlyLeu 346
 b 483 TCTGGTTTA 491
 RESULT 11
 ID AAK44992 standard; DNA; 494 BP.
 GC AAK44992;
 AT AAK44992 (first entry)
 06-NOV-2001 (first entry)
 Human bone marrow expressed single exon probe SEQ ID NO: 19549.
 Human; bone marrow expressed exon; gene expression analysis; probe;
 microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 Homo sapiens.
 WO200157276-A2.
 30-JAN-2001; 2001MO-US000666.
 04-FEB-2000; 2000US-0180312P.
 26-MAY-2000; 2000US-0207456P.
 30-JUN-2000; 2000US-00608408.
 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-489900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human bone marrow.
 XX
 PS Example 4; SEQ ID NO 19549; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention
 CC
 SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,45e-84 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.98% Indels: 0
 DB: Gaps: 0
 US-10-010-050A-2 (1-346) x AAK44992 (1-494)
 QY 184 AlAlaYtRPaYAllySGInAePaengIuThRgLYleTYrTYgluThrTPaenValys 203
 Db 3 GCAAAGGGGTGAACGACGCAATGAACAGCAATTTATATGACATGCAATGTAAAA 62
 QY 204 AlASeRProGluYsgLYAlaGIuThrTPheASerTYrASpCYsSerLYsPheVal 223
 Db 63 GCCAGCCCAAAAAGGGGGGCGAGACACTGGTTGATTCCTACAGCTGTTCCTCAAAATTTG 122
 QY 224 LeuArGrThRPeAenLYsLeuAlaGIuPheGlyAlaGIuPheLYsAenlleGIuThRAsn 243
 Db 123 TTAAAGACCTTTAAACAAGTGGCTGAATTTGGAGCAGAGTTCAAGAACATAGAAACCAAC 82
 QY 244 TYrThArGrIlePheLeuTYrSeRgLYgluPProThrTYrleuGlyAenGIuThRSeVal 263
 Db 183 TATACAGAAATATTCTTTACAGTGAGAACTTACTATCTGGGAAATGAAACATCTGTT 242
 QY 264 PheGLYProThRgLYAenLYsThLeuGlyLeuAlalleYsARgPheTYrTYrProPhe 283
 Db 243 TTTGGGCCAACAGAAACAGACCTCTGGTTTACCAATAAAAAATTTTATTTACCCCTTC 302
 QY 284 LysProHisLeuProThRgLYgluPheLeuSerLeuLeuGlnIlePheASpAlaVal 303
 Db 303 AAACCACTTGGCCAACTTAAGAAATTTCTGTGAGTCTCTTGCAGAAATTTTGAAGCAGTG 362
 QY 304 lIeValHisLYsGInAePaengIuThRgLYleTYrTYgluThrTPaenValys 323
 Db 363 ATTGGCAACAACGTTCTATTGTTTATTAATTTGTAATTTGGTTTTCATATGAA 422
 QY 324 PheProPheIleLYsIleThrTYrGluGluIleProLeuProIleARgAsnLYsThRleu 343
 Db 423 TTCCTCTTATTAATAATATGATGAGAAATCCCTTAACTATACGAAACAAACACTC 482
 QY 344 SerGlyLeu 346
 Db 483 TCTGCTTTH 491
 RESULT 12
 ID AAK19040
 ID AAK19040 standard; DNA; 494 BP.

XX AAK19040;
 AC
 XX 05-NOV-2001 (first entry)
 PT
 XX
 DE Human brain expressed single exon probe SEQ ID NO: 19031.
 XX
 XX Human; brain expressed exon; gene expression analysis; probe; microarray;
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
 KW ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200157275-A2.
 PN
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000667.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483446/52.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains.
 XX
 PS Example 4; SEQ ID NO 19031; 650bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention
 CC
 SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,45e-84 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.98% Indels: 0
 DB: Gaps: 0
 US-10-010-050A-2 (1-346) x AAK19040 (1-494)
 QY 184 AlAlaYtRPaYAllySGInAePaengIuThRgLYleTYrTYgluThrTPaenValys 203
 Db 3 GCAAAGGGGTGAACGACGCAATGAACAGCAATTTATATGACATGCAATGTAAAA 62
 QY 204 AlASeRProGluYsgLYAlaGIuThrTPheASerTYrASpCYsSerLYsPheVal 223
 Db 63 GCCAGCCCAAAAAGGGGGGCGAGACACTGGTTGATTCCTACAGCTGTTCCTCAAAATTTG 122
 QY 224 LeuArGrThRPeAenLYsLeuAlaGIuPheGlyAlaGIuPheLYsAenlleGIuThRAsn 243
 Db 123 TTAAAGACCTTTAAACAAGTGGCTGAATTTGGAGCAGAGTTCAAGAACATAGAAACCAAC 182
 QY 244 TYrThArGrIlePheLeuTYrSeRgLYgluPProThrTYrleuGlyAenGIuThRSeVal 263
 Db 183 TATACAGAAATATTCTTTACAGTGAGAACTTACTATCTGGGAAATGAAACATCTGTT 242

Y PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTYrProPhe 283
 264 |||||
 b TTTGGGCCAAGAAACAAGACTCTGTTAGCCATAAAAGATTATTAACCCCTTC 302
 Y 284 LysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaVal 303
 b 303 AAACCAATTTGGCCAACTTAAGAAATTCGTGTAGTCTCTTGCAAATTTTGTGACAGTG 362
 Y 304 IleValHisLysGlnPheTYrLeuPheTYrAsnPheGlyTYrTrpPheLeuProMetLys 323
 b 363 ATTGTGCACAAACAGTCTTATTTGTTTATTAATTTTGAATTTGGTTTAAACCATGAAA 422
 Y 324 PheProPheIleLysIleThrTYrGluGluIleProLeuProIleArgAsnLysThrLeu 343
 b 423 TTCCCTTTATTAATAATACATATGAAAGAAATCCCTTACTATCAGAAACAAACACATC 482
 Y 344 SerGlyLeu 346
 b 483 TCTGGTTTA 491
 RESULT 13
 BS44658
 ID ABS44658 standard; DNA; 494 BP.
 AC ABS44658;
 CX 25-FEB-2003 (first entry)
 XX Human liver single exon probe, SEQ ID No 19648.
 XX Human; single exon nucleic acid probe; liver; cirrhosis;
 XX hyperlipoproteinemia; hyperlipidaemia; hypercholesterolaemia;
 XX coronary heart disease; ss.
 XX Homo sapiens.
 XX WO200157273-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US000664.
 XX 04-FEB-2000; 2000US-0180312P.
 XX 26-MAY-2000; 2000US-0207456P.
 XX 30-JUN-2000; 2000US-00608408.
 XX 03-AUG-2000; 2000US-00632365.
 XX 21-SEP-2000; 2000US-0234687P.
 XX 27-SEP-2000; 2000US-0236359P.
 XX 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DB;
 XX WPI; 2001-488898/53.
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human adult liver.
 XX Claim 4; SEQ ID NO 19648; 658bp; English.

CC sequence information for this patent does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences
 CC
 XX
 XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,45e-84 Length: 494
 Score: 886.00 Matches: 163
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.98% Indels: 0
 DB: Gaps: 0
 US-10-010-050a-2 (1-346) x ABS44658 (1-494)
 QY 184 AlAlysrTrpVallysgInaAspAsnGluThrGlyIleTYrTYrGluThrTrpAsnVallys 203
 Db 3 GCAAAAGTGGTGAAACAGCAACATGAAACAGAAATTTATATGACATGGAATGTAATA 62
 QY 204 AlAsrProGluYrGlyAlaGluThrTrpPheAspSerTYrAspCysSerLysPheVal 223
 Db 63 GCCAGCCCAAGAAAGGGGGCAGACATGGTTGATTCCTACGACTGTTCCAAATTTGTG 122
 QY 224 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn 243
 Db 123 TTTAAGACCTTTAACAAAGTGGCTGAATTGGACAGAGTTCAAGAAACATGAAACCAAC 182
 QY 244 TyrThrArgIlePheLeuTYrSerGlyGluProThrTYrLeuGlyAsnGluThrSerVal 263
 Db 183 TATACAAAGAAATATTTCTTACAGAGGAGAACCTTATCTGGGAAATGAAACATCTGTT 242
 QY 264 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTYrProPhe 283
 Db 243 TTTGGGCCAAGAAACAAGACTCTGTTAGCCATAAAAGATTATTAACCCCTTC 302
 QY 284 LysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaVal 303
 Db 303 AAACCAATTTGGCCAACTTAAGAAATTCGTGTAGTCTCTTGCAAATTTTGTGACAGTG 362
 QY 304 IleValHisLysGlnPheTYrLeuPheTYrAsnPheGlyTYrTrpPheLeuProMetLys 323
 Db 363 ATTGTGCACAAACAGTCTTATTTGTTTATTAATTTTGAATTTGGTTTAAACCATGAAA 422
 QY 324 PheProPheIleLysIleThrTYrGluGluIleProLeuProIleArgAsnLysThrLeu 343
 Db 423 TTCCCTTTATTAATAATACATATGAAAGAAATCCCTTACTATCAGAAACAAACACATC 482
 QY 344 SerGlyLeu 346
 Db 483 TCTGGTTTA 491
 RESULT 14
 ABS19237
 ID ABS19237 standard; DNA; 494 BP.
 XX ABS19237;
 XX 19-AUG-2002 (first entry)
 XX Human genome-derived single exon probe ORF from lung SEQ ID No 19228.
 XX
 XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 XX chronic obstructive pulmonary disease; interstitial lung disease;
 XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
 XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemostasis;
 XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
 XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 XX primary ciliary dyskinesia; pulmonary hypertension;
 XX hyaline membrane disease; open reading frame; ORF.
 XX Homo sapiens.

human blood, kidney, foetal lung, placenta, testes, brain, ovary,
pituitary, retina and colon cDNA libraries.

Claim 1; Page 366; 6A1pp; English.

The present sequence represents an expressed sequence tag (EST), and is a polynucleotide of the invention. The polynucleotides of the invention are all secreted EST sequences isolated from a variety of human tissue sources. The EST sequences and proteins encoded by them are predicted to have useful biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, haemostatic activity/inhibin regulating activity, chemotactic/chemokinetic activity, and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, tumour inhibition activity. The EST sequences are also stated to be useful for gene therapy

Sequence 506 BP; 171 A; 86 C; 119 G; 130 T; 0 U; 0 Other;

Alignment Scores:

Ref. No.:	1.85e-67	Length:	506
Score:	727.00	Matches:	133
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.73%	Indels:	0
gaps:	2	Gaps:	0

US-10-010-050A-2 (1-346) x AAV88347 (1-506)

151	AlaAlaCysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuVal	170
99	GCTGCCCTCTTTTGGAGGAGATGATGATGTCATCGAAGAAATGGGACATTAATT	158
171	GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGluAsp	190
159	CAAGTAGCACTATATCAAGAAACATGTTCAACCAAAATGGCAAGTGGTAAACAGGAC	228
191	AsnGluThrGlyIleTrpGluThrTrpAsnValLysAlaSerProGluLysGlyAla	220
219	AATGAAACAGAAATTATTATGACATGATGATGAAAGCCAGCCGAAAGGGGCA	278
211	GluThrTrpPheAspSerTrpAspCysSerLysPheValIleuArgThrPheAsnLysLeu	230
279	GAGCATGTGTTGATTCCTACGACTGTTCCAAATTGTGTTAAGGACCTTTACCAAGTTG	338
231	AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTrpThrArgIlePheLeuTyr	250
339	GCTGAATTTGGAGCAGAGTTCAAGAACATGAAACCAACTATACAGAAATATTCTTAC	398
251	SerGlyGluProThrTrpLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys	270
399	AGTGAAGAACCTACTTATCTGGGAAATGAAACATCTGTTTGGGCCAAGCAAGAAACAAG	458
271	ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPhe	283
459	ACTCTTGTTAGCCATAAAAAAGATTTATACCCCTTC	497

Search completed: April 24, 2004, 23:26:22
Job time : 646.658 secs

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1030CE010P1.

FEATURES

source

Location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1030Y01"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 86.6%; Score 960.2; DB 9; Length 1201;
Best Local Similarity 94.0%; Pred. No. 1.2e-199; Indels 8; Gaps 3;
Matches 1012; Conservative 17; Mismatches 40;

```

QY      8 GCACGAGGCGAGAGGTAGACACGCGACAGGCGCGGAGATGCGCGGCGCGCGCGG 67
DB      24 GCAGGCTGTGATCCGGTCCGGGATTCGCCGGATCGCGAGATGCGCGGCGCGCGCGG 63
QY      68 GCTGGGAGACGCGCTTCTGCTGCTGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCT 127
DB      84 GCTCGGGGAGCGCGCTTCTGCTGCTGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCT 143
QY      128 GCGTGTGCGCGGCGCTGCGCGGATCCCTCCCGCGCGCATGCGCGGCTGCTGCTGCTGCT 187
DB      144 GCGTGTGCGCGGCGCTGCGCGGATCCCTCCCGCGCGCATGCGCGGCTGCTGCTGCTGCT 203
QY      188 TTGACTTCCGTCGCAAAACCTGATCTCTTATGTCAGCTAAGTATCTTCTGCTCAACT 247
DB      204 TTGACTTCCGTCGCAAAACCTGATCTCTTATGTCAGCTAAGTATCTTCTGCTCAACT 263
QY      248 GGCCTACCTATCCAGTTATGAGAGGATGATGATGATGATGATGATGATGATGATGAT 307
DB      264 GGCCTACCTATCCAGTTATGAGAGGATGATGATGATGATGATGATGATGATGATGAT 323
QY      308 CCACTATGAGGAAATTAATATGAGAGACCTCTGCGGACATGTAATATATGATGATGCC 367
DB      324 CCACTATGAGGAAATTAATATGAGAGACCTCTGCGGACATGTAATATATGATGATGCC 383
QY      368 ATTGATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 427
DB      384 ATTGATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 443
QY      428 CAATCTGCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 487
DB      444 CAATCTGCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 503
QY      488 AATCAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 547
DB      504 AATCAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 563
QY      548 ACATTAAGTCAAGTGAACAATATGAGAAACATGTTCAACCAATGCGAAAGTGGGTG 607
DB      564 ACATTAAGTCAAGTGAACAATATGAGAAACATGTTCAACCAATGCGAAAGTGGGTG 623
QY      608 AAACAGGACATGAAACAGGAAATTTATATGAGACATGATGATGATGATGATGATGATGAT 667
DB      624 AAACAGGACATGAAACAGGAAATTTATATGAGACATGATGATGATGATGATGATGATGAT 683
QY      668 AAGGGGGGAGAGACATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 727
DB      684 AAGGGGGGAGAGACATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 743
QY      728 AACAGTTGCTGTAATTTGAGAGAGAGTTGACAAATATGAAACCAATATCAAGAAATA 787
DB      744 AACAGTTGCTGTAATTTGAGAGAGAGTTGACAAATATGAAACCAATATCAAGAAATA 803
QY      788 TTCTTTTACAGTGAAGACCTACTATCTGAGGAAATGAAACATCTGTTTTTGGGCCCA 847

```

```

DB      804 TTCTTTTACAGTGAAGACCTACTATCTGAGGAAATGAAACATCTGTTTTTGGGCCCA 863
QY      848 GGAACAGACTCTGTTAGCCATAAAAGATTTATATACCTTCAACCAATTTG 907
DB      864 GGAACAGACTCTGTTAGCCATAAAAGATTTATATACCTTCAACCAATTTG 923
QY      908 CCAACTAAGAAATTTGTTAGTCTCTTGCAGAAATTTTATGACAGTATGTGACAA 967
DB      924 CCAACTAAGAAATTTGTTAGTCTCTTGCAGAAATTTTATGACAGTATGTGACAA 983
QY      968 CAGTTCTATTTGTTTATATTTGAATTTGATTTTACATGATGATGATGATGATGAT 1027
DB      984 CAGTTCTATTTGTTTATATTTGAATTTGATTTTACATGATGATGATGATGATGAT 1038
QY      1028 AAATATACATATGAAGAAATCCCTTACCTATACAGAAACAAACACTCTGCTTTA 1084
DB      1039 TAAATACATATGAGAAATCCCTTTT---CYWTYAAATAAATAACCTCTGCTTTTAA 1092

```

RESULT 2

BX331615

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 923)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f. For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DB009AB010P1&cluster=7238.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DB009AB010P1.

FEATURES

source

1. 923
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB009Y001"
/issue_type="NEUROBLASTOMA COT 10-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 74.9%; Score 811.6; DB 13; Length 923;
Best Local Similarity 96.8%; Pred. No. 4.2e-167; Indels 0; Gaps 0;
Matches 811; Conservative 9; Mismatches 18;

```

QY      14 GGGCAGAGGTAGACAGAGGCGCGGAGATGCGCGGCGCGCGCGGCTGG 73
DB      85 GGCAGAGAGTGAACAGCGACAGGCGCGGAGATGCGCGGCGCGCGCGGCTGG 144
QY      74 GAGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 133
DB      145 GAGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 204

```

```

134 TCCCGGGTCTCGGAGATCCCTCCGCGCCACTGCGCGGTCCTTACAAAGCGCTTTGAC 193
205 TCCCGGGTCTCGGAGATCCCTCCGCGCCACTGCGCGGTCCTTACAAAGCGCTTTGAC 264
194 TTCCTGTCACAAACCTGATCTTATTTGCAAGTAACTATCTTCTGTCACACTGGCTTA 253
265 TTCCTGTCACAAACCTGATCTTATTTGCAAGTAACTATCTTCTGTCACACTGGCTTA 324
254 CCTATCCCGATTTATGAGAGGTGATGATGACATTTAACTTTTTCATTAAGACCCCACTA 313
325 CCTATCCCGATTTATGAGAGGTGATGATGACATTTAACTTTTTCATTAAGACCCCACTA 384
314 TGGGAATTTAAATATGAGAGCTCTCTGAGACACTTGAATAATATGATGATGATGATGAT 373
385 TGGGAATTTAAATATGAGAGCTCTCTGAGACACTTGAATAATATGATGATGATGATGAT 444
374 TTCAGAAATGATCATTAACTGAGCAAGAACTACAAATGATGATGATGATGATGATGAT 433
445 TTCAGAAATGATCATTAACTGAGCAAGAACTACAAATGATGATGATGATGATGATGAT 504
434 GCGAATCTGATCATTCCCGATCTCCGACCTGAATGATGATGATGATGATGATGATGAT 493
505 GCGAATCTGATCATTCCCGATCTCCGACCTGAATGATGATGATGATGATGATGATGAT 564
494 GCGCTGCTGCTCTTTTGAAGGATGATGATGATGATGATGATGATGATGATGATGAT 553
565 GCGCTGCTGCTCTTTTGAAGGATGATGATGATGATGATGATGATGATGATGATGAT 624
554 GTTCAAGTACAACTATATGAGAAACATGTTCAACCAATGATGATGATGATGATGATGAT 613
625 GTTCAAGTACAACTATATGAGAAACATGTTCAACCAATGATGATGATGATGATGATGAT 684
614 GACAAATGAAACAGAAATTTATGAGCAATGATGATGATGATGATGATGATGATGATGAT 673
685 GACAAATGAAACAGAAATTTATGAGCAATGATGATGATGATGATGATGATGATGATGAT 744
674 GCAAGAGCATGTTGATCTCTACAGCTGTTCCAAATTTGTTTGAAGACCTTTTACAAAG 713
745 GCAAGAGCATGTTGATCTCTACAGCTGTTCCAAATTTGTTTGAAGACCTTTTACAAAG 804
734 TTGCTGATTTTGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 793
805 TTGCTGATTTTGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 864
794 TACAGTGAAGAACTTACTATCTGGAATGAAATGATGATGATGATGATGATGATGATGAT 851
865 TACAGTGAAGAACTTACTATCTGGAATGAAATGATGATGATGATGATGATGATGATGAT 922

```

```

RESULT 3          1077 bp      mRNA      linear      EST 02-MAY-2003
LOCUS            BX342662 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION      Homo sapiens cDNA clone CS0DL004YP03 5-PRIME, mRNA sequence.
ACCESSION       BX342662
VERSION         BX342662.1 GI:3038103
KEYWORDS
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1077)
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segefe@genoscope.cns.fr; Web : www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f For
more information about this cluster, see

```

```

http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DL004YP03&cluster=7238.f. Contact :
Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DL004YP03.
Location/Qualifiers
1..1077
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL004YP03"
/cell_line="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

```

Query Match          73.1%; Score 792.4; DB 13; Length 1077;
Best Local Similarity 90.0%; Pred. No. 6.7e-163;
Matches 821; Conservative 49; Mismatches 36; Indels 6; Gaps 3;

```

```

14 GGGCAGAGGTAGACAGGCGACAGGCGCGGAGATGCGGCGGCGCGGCGGCGGCGGCGGCGG 72
82 GCGCAGAGGTAGACAGGCGACAGGCGCGGAGATGCGGCGGCGCGGCGGCGGCGGCGGCGG 141
73 GGGACGCGCTTCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 132
142 GGGACGCGCTTCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 201
133 GTCCCGGGTCTCGGAGATCCCTCCGCGCGCACTGGCGGCGGCGGCGGCGGCGGCGGCGG 192
202 GTCCCGGGTCTCGGAGATCCCTCCGCGCGCACTGGCGGCGGCGGCGGCGGCGGCGGCGG 261
193 GTCCCGGGTCTCGGAGATCCCTCCGCGCGCACTGGCGGCGGCGGCGGCGGCGGCGGCGG 252
262 GTCCCGGGTCTCGGAGATCCCTCCGCGCGCACTGGCGGCGGCGGCGGCGGCGGCGGCGG 321
253 ACCTATCCAGTATGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 312
322 ACCTATCCAGTATGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 381
313 ATGGAATTTAAATATGAGAGCTCTGAGCACTTGAATTTATGATGATGATGATGATGATGAT 372
382 ATGGAATTTAAATATGAGAGCTCTGAGCACTTGAATTTATGATGATGATGATGATGATGAT 441
373 ATGGAATTTAAATATGAGAGCTCTGAGCACTTGAATTTATGATGATGATGATGATGATGAT 432
442 ATGGAATTTAAATATGAGAGCTCTGAGCACTTGAATTTATGATGATGATGATGATGATGAT 501
433 TGGCACTGATGATTTCCCATCTTCCGACCTGGAATGATGATGATGATGATGATGATGATGAT 492
502 TGGCACTGATGATTTCCCATCTTCCGACCTGGAATGATGATGATGATGATGATGATGATGAT 561
493 AGGCGCTGCTGCTTTTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 552
562 AGGCGCTGCTGCTTTTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
553 AGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 612
622 AGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 681
613 GGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 672
682 GGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 741
673 GGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 732
742 GGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 801

```

733 GTTGCTGAATTTGAGCAGAGTTCAAGACATAGAAACCACTATACAGAAATATTTCT 752
 802 ATTGGCTGAATTTGAGCAGAGTTCAAGACATAGAAACCACTATACAGAAATATTTCT 861
 793 TTACAGTGGAGACCTACTATCTGGGAATGATAACATCTGTTTGGGCCACAGAGAA 852
 862 TTWAGTGGAGAR-CTACTTATCTGGGAATGATAACATCTGTTTGGGCCACAGAGAC 920
 853 CAAGACTCTGTTAGCCATPAAAAAGATTATTAATCCCTTCAACCACTTTGCCAAC 922
 921 MAACCTCTTGTGTGCTAAAGATTATTA-----SCYTCMARCCMMTTGSCAMHA 976
 913 TAAAGATTTCT 924
 977 AAAATKTTTT 988

RESULT 4
 BI090566
 LOCUS
 DEFINITION
 60285673P1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996891 5',
 mRNA sequence.

ACCESSION
 BI090566.1 GI:14508896
 EST.
 KEYWORDS
 Homo sapiens (human)
 SOURCE
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 (bases 1 to 869)
 NIH-MGC http://mgs.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE
 Unpublished (1999)
 JOURNAL
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLNI at:
 http://image.llnl.gov
 Plate: LHAM1023 row: a column: 20
 High quality sequence stop: 843.

FEATURES
 source
 1..869
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4996891"
 /cell_line="MGC36"
 /lab_host="DH10B"
 /clone_id="NIH_MGC_10"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: Not";
 Site 2: SalI; Cloned unidirectional. Primer: Oligo dT.
 Average insert size 1.5 kb. Library prepared by Life
 Technologies."

ORIGIN

Query Match 69.1%; Score 749.4; DB 12; Length 869;
 Best Local Similarity 97.1%; Pred. No. 1.7e-153;
 Matches 827; Conservative 0; Mismatches 17; Indels 8; Gaps 6;
 14 GGGCAGAGGTAGACACGGCAGAGGCGCGAGATGGCGCGCGCGCGCGCGCGCG 73
 19 GCGCAGAGGTAGACACGGCAGAGGCGCGCGAGATGGCGCGCGCGCGCGCGCGCG 78
 74 GAGAGCGCTTCTGCTGCTGCGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 133
 79 GAGAGCGCTTCTGCTGCTGCGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 138
 134 TCCCGGCTCTCGGGCATCCCTCCCGGCGCACTGGCGGTGCCCTACAGCGCTTTGAC 193

Db 139 TCCCGGCTCTCGGGCATCCCTCCCGGCGCACTGGCGGTGCCCTACAGCGCTTTGAC 198
 194 TTCCGTCAAAACCTGATCCTTATTTGTCAGCTAGATATCTTGTGCACTGCTCA 253
 199 TTCCGTCAAAACCTGATCCTTATTTGTCAGCTAGATATCTTGTGCACTGCTCA 258
 254 CCTATCCAGTTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATG 313
 259 CCTATCCAGTTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATG 318
 314 TGGGAATTTAAATATGAGACCTCTGCGGACCTTGGAAATTTATGATGATGATG 373
 319 TGGGAATTTAAATATGAGACCTCTGCGGACCTTGGAAATTTATGATGATGATG 378
 374 TTGAGAGTACATTAATGAGAGACCTGAGACCTGAGACCTGAGACCTGAGACCT 433
 379 TTGAGAGTACATTAATGAGAGACCTGAGACCTGAGACCTGAGACCTGAGACCT 438
 434 GGCACATGATATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGAT 493
 439 GGCACATGATATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGAT 498
 494 GGCCTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATG 553
 499 GGCCTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATG 558
 554 GTTCAGATGACATATATACAGAAACATGTTCAACCAATGGCAATGGTGGTGAAC 613
 559 GTTCAGATGACATATATACAGAAACATGTTCAACCAATGGCAATGGTGGTGAAC 617
 614 GACATATGAAACAGAAATTTATATGAGCATGATGATGATGATGATGATGATGAT 671
 618 GACATATGAAACAGAAATTTATATGAGCATGATGATGATGATGATGATGATGAT 677
 672 GGGCAGAGCATGTTTATGTTATCTTCAAGCTGTTCCAAATTTGTGTAAGACCTT 731
 678 GGCACATGATGATGTTTATGTTATCTTCAAGCTGTTCCAAATTTGTGTAAGAC 737
 732 AGTTGGCTGAA-TTGGAGCAGAGTTCAAGAA-TAGAAACCACTATACAGAAAT 789
 738 AGTTGGCTGAACTTGGAGCAGAGTTCAAGAACTAGAAACCACTATACAGAAAT 797
 790 TCTTTACAGTGG-AGAACCTTATCTGCGGAAA--TGAAACATCTGTTTGGCGAAC 846
 798 TCTTTACAGTGGAAAGCTTATCTGCGGAAAATGAAACATCTGTTTGGCGAAC 857
 847 AGGAAACAGAC 858
 858 AGGAAACAGAC 869

RESULT 5
 CA488543
 LOCUS
 DEFINITION
 AGENCOURT_10808864 MAPCL Homo sapiens cDNA clone IMAGE:6720309 5',
 mRNA sequence.
 ACCESSION
 CA488543
 VERSION
 CA488543.1 GI:24950702
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 1 (bases 1 to 924)
 NIH-MGC http://mgs.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE
 Unpublished (1999)
 JOURNAL
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 CDNA Library Preparation: Invitrogen Corp.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLMT at:
<http://image.llnl.gov>
 Plate: LLMT4279 row: j column: 21
 High quality sequence stop: 637.
 Location/Qualifiers

FEATURES

source

1. 924
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6720309"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, hTERT-HME1, LNCaP"
 /lab_host="EMD10B"
 /clone_lib="MAPcl"
 /note="Vector: pCMV.SPORT6, Site 1: EcoRV, Site 2: Not I; Subcloned with brain, liver, lung, kidney and muscle. Directionally cloned. Priming method: oligo-dt. Average insert size: 1800 bp. Library amplification: 26,000 fold. Kristi A. Egland, James J. Vincent, Robert Strausberg, Bungkok Lee & Ira Pastan: Discovery of new breast cancer genes encoding membrane and secreted proteins. Manuscript submitted."

ORIGIN

Query Match 68.6%; Score 743.4; DB 14; Length 924;
 Best Local Similarity 96.7%; Pred. No. 3.6e-152;
 Matches 759; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

32 GCACAGGCGCGCGAGATGCGGCGGCGGCGGCGGCTCGGGACGCGCTTCTGTGTC 95.
 1 GGACAGGCGCGCGAGATGCGGCGGCGGCGGCGGCTCGGGACGCGCTTCTGTGTC 60

92 TGGGCGCTGGGCGCTTGGCTGGCGGCTGGCTGGCGGCTGGCGGCTGGCGGCTGG 151
 61 TGGGCGCTGGGCGCTTGGCTGGCGGCTGGCTGGCGGCTGGCGGCTGGCGGCTGG 120

152 CCTCCGCGGCGCGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 211
 121 CCTCCGCGGCGCGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 180

212 CCTATTGTCAAGCTAAGTATATCTTGTGTCACTGGCTCACTTATCCGATATGAG 271
 181 CCTATTGTCAAGCTAAGTATATCTTGTGTCACTGGCTCACTTATCCGATATGAG 240

272 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 331
 241 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

332 GACCTCCGCGGCGCGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 391
 301 GACCTCCGCGGCGCGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 360

392 GGCAGAACTACACAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 451
 361 GGCAGAACTACACAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420

452 CATCTCCGCGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 511
 421 CATCTCCGCGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 480

512 GAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 571
 481 GAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540

572 TCAGAAACATGTTTCAACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 631
 541 TCAGAAACATGTTTCAACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 600

632 TATTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 691
 601 TATTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660

QY 692 TCTACGACTGTTCACAAATTTGTGTTAAGACCTTTACACAGTGTGTAATTTGAGCA 751
 DB 661 TCTACGACTGTTCACAAATTTGTGTTAAGACCTTTACACAGTGTGTAATTTGAGCA 720

QY 752 GAGTTCAGAACATGAAACCAACTATACAGAAATTTCTTTACAGTGAAGAACTACT 811
 DB 721 GAGTTCAGAACATGAAACCAACTATACAGAAATTTCTTTACAGTGAAGAAACC 780

QY 812 TATCT 816
 DB 781 TTAAT 785

RESULT 6
 BI223533
 LOCUS
 DEFINITION
 602941923F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:510503 5', mRNA sequence.
 BI223533
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 787)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: sgabs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLMT at:
<http://image.llnl.gov>
 Plate: LLMT1253 row: n column: 06
 High quality sequence stop: 762.
 Location/Qualifiers

FEATURES

source

1. 787
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:510503"
 /tissue_type="cervical carcinoma cell line"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 12"
 /note="Organ: cervix; Vector: pCMV.SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.4 kb. Library prepared by Life Technologies."

ORIGIN

Query Match 65.8%; Score 712.8; DB 12; Length 787;
 Best Local Similarity 98.5%; Pred. No. 1.8e-145;
 Matches 772; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

22 GGTAGACACGCGACAGCGCGCGAGATGCGCGGCGCGCGCGCGCGCGCGCGCGCG 81
 1 GGTAGACACGCGACAGCGCGCGAGATGCGCGGCGCGCGCGCGCGCGCGCGCGCG 60

82 TTCTGTGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 141
 61 TTCTGTGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGG 120

142 CTCGGGCAATCCCTCCGCGCGCACTGCGCGGCTGCGCTTCAAGCGCTTTGACTTCGTC 201
 121 CTCGGGCAATCCCTCCGCGCGCACTGCGCGGCTGCGCTTCAAGCGCTTTGACTTCGTC 180

202 AAAACCTGATCTTATTTGCAAGTATATCTTGTGTCACTGCTCACTTATCC 261

[illegible][illegible]

582 CAAGTGGCCGAGTGGGTGAAGACGACGACATGAACTGGATTTATATGACATGGAC 641
 649 TGTAAAGCCAGCCAGAAAAGG3G3GAGAGACATGTTTATTTCTTACGACTCTTCCAA 708
 642 GGTCCGAGCCGCCAGGAGCAAGG3G3GAGAGACATGTTTATTTCTTACGACTCTTCCAA 701
 709 ATTGTGTAAAGACCTTTAAAGATTGGCTGAATTTGGAGAGAGTCAAGACATAGA 768
 702 TTTTGTCTTAAGGACATATTAAGAAATTTGCTGAATTTGAACAGATTCCAGGAATAGA 761
 769 AACCAACTATACAGAAATATTTCTTACAGTGAAGACCTATCTCTGGAAATGAAC 828
 762 AACCAACTATACAGAAATATTTCTTACAGTGAAGACCTATCTCTGGAAATGAAC 821
 829 ATCTGTTTTGGGCCCAACAGAAAGACCTCTGTTTATTTGACCAATTAATTTATTA 838
 822 ATCTATTTTGGGCCCAACAGAAAGACCTCTGTTTATTTGACCAATTAATTTATTA 831
 889 CCCCTTCAACACATTTTCCCACTTAAGAAATTTGCTGAGTCTCTGCAATTTTGA 948
 882 CCCCTTCAACACATTTTCCCACTTAAGAAATTTGCTGAGTCTCTGCAATTTTGA 941
 949 TGAAGATGTTGTCACAAACAGTCTATTTGTTTATTAATTTGAAATTTGTTTACC 1008
 942 TACAGTATTTATACAGACAGTCTTACTTGTTTATTAATTTGAGTATTTGTTTACC 1001
 1009 TATGAATTCCTTTTATTAATTAATTAATTAATTAATTTGAAATTTGTTTACC 1068
 1002 AATGAACACCCCTTTGTCAAAATTAATTAATTAATTAATTTGAAATTTGTTTACC 1061
 1069 AACACTCTCTGTTT 1083
 1062 AACATTATCCGACTT 1076

RESULT 9
 AK085741 2433 bp mRNA linear HTC 20-SEP-2003
 LOCUS AK085741
 DEFINITION Mus musculus 10 days lactation, adult female mammary gland cDNA.
 RIKEN full-length enriched library, clone: D73003P03
 product: similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLNS
 PROTEIN) [Homo sapiens], full insert sequence.

ACCESSION AK085741
 VERSION AK085741.1 GI:26102938
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimura, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujisawa, S., Inoue, K., Togawa, Y., Iizaka, M., Ohara, E., Matsui, M.,
 Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawa, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE
 AUTHORS 4
 TITLE The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 JOURNAL Functional annotation of a full-length mouse cDNA collection
 REFERENCE Nature 409, 685-690 (2001)
 AUTHORS 5
 TITLE The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 JOURNAL Analysis of the mouse transcriptome based on functional annotation
 REFERENCE of 60,770 full-length cDNAs
 AUTHORS Nature 420, 563-573 (2002)

6 (bases 1 to 2433)
 Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hixomoto, K., Hirooka, T., Hirozane, T.,
 Hori, F., Imotani, K., Itoh, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, R., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Tanaka, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in RIKEN
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in RIKEN contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers

FEATURES
 source
 1..2433
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:D73003P03"
 /db_xref="MGJ:2422937"
 /db_xref="taxon:10090"
 /clone="D73003P03"
 /sex="female"
 /tissue type="mammary gland"
 /clone.lib="RIKEN full-length enriched mouse cDNA library"
 /dev stage="10 days lactation, adult"

misc_feature
 1..2433
 /note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5
 (CLNS PROTEIN) [Homo sapiens] (SWISSPROT:O1075503, evidence:
 FASTV, 73.9%ID, 86.7%length, match=1041)"

ORIGIN
 Query Match 62.6% Score 678.6; DB 11; Length 2433;
 Best Local Similarity 79.2%; Pred. No. 5.7e-118;
 Matches 820; Conservative 0; Mismatches 209; Indels 6; Gaps 1;
 49 GCGGGGGGCGGCGGCGGCTCGGGGAGCGGCTTCTGCTGCTGGCGGCTGCTGCT 108
 125 GCGGGGAGTCTCGCGCGGCGGCGGCTGAGCGGCGGCGGCGGCTGCGCTGCG 184
 109 TTGGCTCGGCGGCTTCCGCGGCTGCTCCGCGGCTTCCGCGGCGGCGGCGGCTG 168
 185 GGTGCTCGGCGGCTGCGGCGGCTGCTCGGCGGCTGCGGCGGCTGCGGCGGCTG 238

```

169 GCCGCTGCTTACAGAGCGTTGACTTCGTCACAAACCTGATTCCTATTGTCAAGCTAA 223
b 239 GCGGCGCCCTACAGAGCGTTCTCTTTCGTCGGAAGACAGATCCCTACGTCAAGCTAA 293
Y 229 GTATACCTTCTGTCACAGCTGCTACCTATCCAGTTATGAGAGGTGATGATCACTTGA 288
Y 299 GTATACCTTCTGTCACAGCTGCTACCTATCCAGTTATGAGAGGTGATGATCACTTGA 358
Y 289 AGTTTTCATTAACAGAGCGTTGAGTTTAAATGAGAGCGTCTGAGAGCTT 348
b 359 GGTCTTACGACTACAGAGCGTTGAGTTTAAATGAGAGCGTCTGAGAGCTT 418
Y 349 GAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
b 419 TAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 478
Y 409 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 468
b 479 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 538
Y 469 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 528
b 539 GAGGCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
Y 529 TCATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538
b 539 ACATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
Y 589 CCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648
b 659 CAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 718
Y 649 TGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 708
b 719 GGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778
Y 709 ATTTGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768
b 779 TTTTGTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838
Y 769 AACCACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 828
b 839 AACCACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
Y 829 ATCTGTTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 888
b 899 ATCTGTTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 958
Y 889 CCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 948
b 959 CCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018
Y 949 TGCAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1008
b 1019 TACAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078
Y 1009 TATGAAATTCCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1068
b 1079 AATGAAATTCCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1138
Y 1069 AACACTCTCTGCTT 1083
b 1139 AACACTCTCTGCTT 1153

```

RESULT 10
 AL571805/c 1201 bp mRNA linear EST 31-MAY-2003
 LOCUS AL571805 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 DEFINITION clone CS0D1030YJ01 3-PRIME, mRNA sequence.
 ACCESSION AL571805
 VERSION AL571805.2 GI:31293196

KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1201)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 16, 2001 this sequence version replaced gi:12929467.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 7238.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0D1030CE01NP1&cluster=7238.f. Contact :
 Feng Liang Email : fliang@life.techn.com URL :
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0D1030CE01NP1.
 Location/Qualifiers
 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1030YJ01"
 /issue_type="PLACENTA COT 25-NORMALIZED"
 /clone_1ib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo (dt)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
 source

ORIGIN
 Query Match 62.2%; Score 674.2; DB 9; Length 1201;
 Best Local Similarity 97.8%; Pred. No. 5.1e-137;
 Matches 704; Conservative 9; Mismatches 4; Indels 3; Gaps 3;

```

QY 366 CCATTGATTCAGAAATGATCACTTACCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 425
DB 1075 SCATGATTCAGAAATGATCACTTACCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1018
QY 426 TCCAACTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
DB 1017 TCCAACTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 958
QY 486 GTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
DB 957 GTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
QY 545 GGCAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
DB 897 GGCAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838
QY 605 GTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 664
DB 837 GTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778
QY 665 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 724
DB 777 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 718
QY 725 TTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 784
DB 717 TTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658
QY 785 ATATTCCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 844
DB 657 ATATTCCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
QY 845 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 904

```

Db 597 ACAGAAACAAAGACCTTGGTTAGCCATAAAATTTATTTACCCCTTAACACCAT 538
 Qy 905 TTGGCAACTGAAGAAATTTCTGTAGTCTCTGCAAAATTTTGTATGACAGTATGTGCAC 934
 Db 537 TTGGCAACTGAAGAAATTTCTGTAGTCTCTGCAAAATTTTGTATGACAGTATGTGCAC 478
 Qy 965 AACAGTTCTATTTGTTTATTAATTTTGAATATGTGTTTACCTATGAAATTCCTTTT 1024
 Db 477 AACAGTTCTATTTGTTTATTAATTTTGAATATGTGTTTACCTATGAAATTCCTTTT 418
 Qy 1025 ATTTAAATATACATATGAGAAATTCCTTACCTATGAGAAACAAACCTCTGTGTTTA 1084
 Db 417 ATTTAAATATACATATGAGAAATTCCTTACCTATGAGAAACAAACCTCTGTGTTTA 358
 RESULT 11
 BE873363 931 bp mRNA linear EST 20-OCT-2000
 LOCUS 601450425F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3854119 5',
 DEFINITION mRNA sequence.
 BE873363
 ACCESSION BE873363.1 GI:10322139
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 931)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LHM9579 row: b column: 08
 High quality sequence stop: 662.
 Location/Qualifiers
 1..931
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3854119"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_65"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally. Primer: oligo 3T.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."
 ORIGIN
 Query Match 62.1%; Score 673; DB 10; Length 931;
 Best Local Similarity 93.0%; Pred. No. 9.3e-137;
 Matches 727; Conservative 0; Mismatches 51; Indels 4; Gaps 2;
 Qy 274 TGATGATGACATGTAAGTTTTCGATTCACAGCCCAAGTATGGAATTTAATATGAGA 333
 Db 1 TGATGATGACATGTAAGTTTTCGATTCACAGCCCAAGTATGGAATTTAATATGAGA 60
 Qy 334 CCTCTGGAGACCTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATG 393
 Db 61 CCTCTGGAGACCTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATG 120
 Qy 394 CAAGAACTACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 453
 Db 121 CAAGAACTACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180

Qy 454 TCTCCGACCTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 513
 Db 181 TCTCCGACCTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
 Qy 514 GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 573
 Db 241 GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 Qy 574 AGGAAACATGTTTCAACCAATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 633
 Db 301 AGGAAACATGTTTCAACCAATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 360
 Qy 634 TTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 693
 Db 361 TTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 Qy 694 CTACGACTGTTCCAAATTTGTTTAAAGACCTTTTAAACAATTTGCTGATTTGAGCAGA 753
 Db 421 CTACGACTGTTCCAAATTTGTTTAAAGACCTTTTAAACAATTTGCTGATTTGAGCAGA 480
 Qy 754 GTTCAAGAACATGAGAACCAACTATACAGAAATTTTCTTTACAGTGGAGAACCTACTTA 813
 Db 481 GTTCAAGAACATGAGAACCAACTATACAGAAATTTTCTTTACAGTGGAGAACCTACTTA 540
 Qy 814 TCTGGGAATGAAACATCTGTTTGGGCAACAGGAAACAGACTCTTGTGTAGCCAT 873
 Db 541 TCTGGGAATGAAACATCTGTTTGGGCAACAGGAAACAGACTCTTGTGTAGCCAT 600
 Qy 874 AAAAAGATTTTATACCCCTTCAACCACTTTGCAATTTGCAATTTCTGTAGTCT 933
 Db 601 AAAAAGATTTTATACCCCTTCAACCACTTTGCAATTTGCAATTTCTGTAGTCT 660
 Qy 934 CTTGCAATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 993
 Db 661 CTTGCAATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 716
 Qy 994 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1053
 Db 717 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 776
 Qy 1054 AC 1055
 Db 777 TC 778
 RESULT 12
 BU621797/c 689 bp mRNA linear EST 23-SEP-2002
 LOCUS BU621797
 DEFINITION UT-H-FL1-bgd-1-07-0-UT.s1 NCI CGAP FL1 Homo sapiens cDNA clone
 BU621797
 UT-H-FL1-bgd-1-07-0-UT 3', mRNA sequence.
 BU621797.1 GI:23288012
 ACCESSION
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 689)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: James Martin
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 Seq primer: M13 FORWARD
 POLI=Yes.

FEATURES
source

Location/Qualifiers
1. 689
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FLI-bdp-1-07-0-UI"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_11b="NCI CGAP FLI"
/note="Organ: Chondrosarcoma; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: Bcor I; Site 2: Not I; NCI CGAP FLI is a normalized cDNA library derived from a pool of mRNA obtained from 4 cell lines from grade III chondrosarcoma tissues. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GAGGTCCGTG. The cell lines were provided by Dr. James Martin from the University of Iowa.
TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix
TAG_LIB=UI-H-FLI
TAG_SEQ=GAGGTCCGTG"

IRGIN

Query Match 61.9%; Score 670.6; DB 13; Length 689;
Best Local Similarity 99.4%; Pred. No. 3,1e-136;
Matches 673; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

370 TGAATTCGAAAGTACATTACTGCGAAGACCTACCAATGGAATGGTATGTTTCCCA 429
369 TGAATTCGAAAGTACATTACTGCGAAGACCTACCAATGGAATGGTATGTTTCCCA 630
430 ACTTGGCAAGCTGTCATTTCCCAATCTCCGACCTGGAATGGATGCGCTTCTGGTGA 489
629 ACTTGGCAAGCTGTCATTTCCCAATCTCCGACCTGGAATGGATGCGCTTCTGGTGA 570
490 TCAAGGCGCTGCTCTTTTGAAGGAATGATGATGTTCACTGGAAGAAATGGGAC 549
569 TCAAGGCGCTGCTCTTTTGAAGGAATGATGATGTTCACTGGAAGAAATGGGAC 510
550 ATTAGTTCAAGTAGCACTATATTCAGAAACATGTTCAACCAATGGCAAAAGTGGTAA 639
509 ATTAGTTCAAGTAGCACTATATTCAGAAACATGTTCAACCAATGGCAAAAGTGGTAA 430
610 ACAGACATGAAAGAAAGAAATTTATGAGACATGATGATGAAAGCCAGCCAGAAAA 659
449 ACAGACATGAAAGAAAGAAATTTTGAAGACATGATGATGAAAGCCAGCCAGAAAA 390
670 GGGGGCAGAGCATGTTGATTCCTACGACTGTTCCAAATTTGTTTAAAGACCTTTAA 729
389 GGGGGCAGAGCATGTTGATTCCTACGACTGTTCCAAATTTGTTTAAAGACCTTTAA 330
730 CAAGTGGCTGAATTTGAGCAGAGTTCAAGAACATGAAACCACTATACAAAGAAATTT 789
329 CAAGTGGCTGAATTTGAGCAGAGTTCAAGAACATGAAACCACTATACAAAGAAATTT 270
790 TCTTTCACTGAGAACCTACTATATGAGGAATGAAACATGTTTGGGCAACAG 849
269 TCTTTCACTGAGAACCTACTATATGAGGAATGAAACATGTTTGGGCAACAG 210
850 AAACAAGACTCTTGTGTAGCCATAAAAGATTTTATACCCCTTCAACCACTTTGCC 509
209 AAACAAGACTCTTGTGTAGCCATAAAAGATTTTATACCCCTTCAACCACTTTGCC 150
910 AACTAAAGATTTCTGTGAGTCTCTTGCAAATTTTGAATGCGATGATTTGTGACAAACA 569

```

Db 149 AACTAAAGATTTCTGTGAGTCTCTTGCAAATTTTGAATGCGATGATTTGTGACAAACA 90

Qy 970 GTTCTATTTGTTTAAATTTGAAATATGTTTAACTATGAAATTCCTTTATTTA 1029

Db 89 GTTCTATTTGTTTAAATTTGAAATATGTTTAACTATGAAATTCCTTTATTTA 30

Qy 1030 AATAACATATGAAAGAA 1046

Db 29 AATAACATTTGAAAGAA 13

RESULT 13
AK032293
LOCUS DEFINITION
AK032293 2247 bp mRNA linear HTC 18-SEP-2003
Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone 643051BP20 product: similar to CERIOD-1/POFUSCINOSIS NEURONAL PROTEIN 5 (CLN5 PROTEIN) [Homo sapiens], full insert sequence.
AK032293
ACCESSION
AK032293.1 GI:26328112
VERSION
AK032293.1
KEYWORDS
HTC; GAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
AUTHORS
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
MEDLINE
PUBMED
99279253
10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugihara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtriction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE
PUBMED
20499374
11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, T., Nishi, K., Kibumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL
MEDLINE
PUBMED
20530913
11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL
MEDLINE
PUBMED
11076861
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL
MEDLINE
PUBMED
12000000
6 (bases 1 to 2247)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hitozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kuritani, C., Matsuyama, T., Miyazaki, A., Murata, N., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shitaki, T.,

TITLE

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-research@riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://phantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES

source

```
1..2247
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:6430516P20"
/db_xref="MGI:2395762"
/db_xref="taxon:10090"
/clone="6430516P20"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
<1..971
/note="unnamed protein product; putative
similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5
PROTEIN) [Homo sapiens] (SWISSPROT |075503, evidence:
FASTA, 73.9%ID, 86.7%length, match=1041)."
/codon_start=3
/protein_id="BAC27797.1"
/db_xref="GI:26328113"
/translation="LGIATLIGASPTSGQRPVYKHFSEPRTPDYCAKRTYCPG
SPIPKMDNDVIEVLRLQAPIMEFKYGDLDGHRKLMHDVGFSTLTGKTYLIEWEL
FOJGNCSTPHLRDCKSPKPCWNGOAGCFEFGIDDKMKHGLSVALLISGNTFNFLA
EMVKQDNRETCIYETETVTPAGPGOAGTPEESYDGNFVRYTKLAERFTEPKLET
NRYKIFLYSGEPIYLGNETSIFGPKGKNTALAKKXGPGRPRLSTKDFLMMFLKIF
DVTIIRHOFPLFYFEXWFLPMKPPFKTYEBETPLPRTHTTTDL"
```

CDS

```
137 CGGGCTCGGGGCAATCCCTCCCGGCGGCACTGCGCGGTGCCCTTCAAGCGCTTGACTTC 196
21 CTGGGCGGTGCCCGGAGCTGTGGGCAACGCTGGCGGTCATACCAAGCACTTCTCTTTC 30
197 CGTCCAAAACCTGATCTTATTGCAAGCTAAGTATCTTCTGCAACGCTGCACT 256
81 CGTCCGAAGACAGATCCCTACTGTCACCTAAGTATCTTCTGCTTCCGAGCTGCC 140
257 ATCCGAGTTATGGAGGCTGATGATGACATGAAGTTTTCATTACAGCCCAAGTATG 316
141 ATCCGAGTTATGGAGGCTGATGATGACATGAAGTTTTCATTACAGCCCAAGTATG 200
317 GAATTTAAATATGAGACCTCTCTGGACACTTGAATTTATGATGATGATGATGATG 376
201 GAATTTAAATATGAGACCTCTCTGGACACTTGAATTTATGATGATGATGATGATG 260
377 AGAAGTACATTTACTGCGCAAGCACTACCAATGGAATGGATGATGATGATGATGATG 436
261 AGGACACACACTGACAGGCAAGCACTACCAATGGAATGGATGATGATGATGATGATG 320
437 AACTGTACATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGATG 496
```

ORIGIN

Query Match 61.7%; Score 668.6; DB 11; Length 2247;
Best Local Similarity 81.6%; Pred. No. 8.8e-136;
Matches 773; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 137 CGGGCTCGGGGCAATCCCTCCCGGCGGCACTGCGCGGTGCCCTTCAAGCGCTTGACTTC 196
DB 21 CTGGGCGGTGCCCGGAGCTGTGGGCAACGCTGGCGGTCATACCAAGCACTTCTCTTTC 30
QY 197 CGTCCAAAACCTGATCTTATTGCAAGCTAAGTATCTTCTGCAACGCTGCACT 256
DB 81 CGTCCGAAGACAGATCCCTACTGTCACCTAAGTATCTTCTGCTTCCGAGCTGCC 140
QY 257 ATCCGAGTTATGGAGGCTGATGATGACATGAAGTTTTCATTACAGCCCAAGTATG 316
DB 141 ATCCGAGTTATGGAGGCTGATGATGACATGAAGTTTTCATTACAGCCCAAGTATG 200
QY 317 GAATTTAAATATGAGACCTCTCTGGACACTTGAATTTATGATGATGATGATGATG 376
DB 201 GAATTTAAATATGAGACCTCTCTGGACACTTGAATTTATGATGATGATGATGATG 260
QY 377 AGAAGTACATTTACTGCGCAAGCACTACCAATGGAATGGATGATGATGATGATGATG 436
DB 261 AGGACACACACTGACAGGCAAGCACTACCAATGGAATGGATGATGATGATGATGATG 320
QY 437 AACTGTACATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGATG 496

DB 321 AACTGTACATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGATGATG 380
QY 497 GCGGCCGCTTTTGTGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 556
DB 381 GCAAGCTGCTTTTGTGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 440
QY 557 CAACTAGCAATATATGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 616
DB 441 GTCTTTCACACATATCTCCGAAACATTTTAAACAAAGTGGCGGATGATGATGATGATG 500
QY 617 AATGAAACAGGAATTTATATGATGATGATGATGATGATGATGATGATGATGATGATG 676
DB 501 AATGAAACAGGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 560
QY 677 GAGACATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 736
DB 561 CAGAGTGTGCTGAGTCTTACGATGATGATGATGATGATGATGATGATGATGATGATG 620
QY 737 GCTGATTTGAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 796
DB 621 GCTGATTTGAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 680
QY 797 AGTGAAGACCTACTTATCTGAGGAATGATGATGATGATGATGATGATGATGATGATG 856
DB 681 AGTGAAGACCTACTTATCTGAGGAATGATGATGATGATGATGATGATGATGATGATG 740
QY 857 ACTCTTGTGTTAGCATAAAGATTTATTAACCTCTTCAACCAATTTGCCAATTA 916
DB 741 ACTCTTGTGTTAGCATAAAGATTTATTAACCTCTTCAACCAATTTGCCAATTA 800
QY 917 GAATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 976
DB 801 GAATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 860
QY 977 TTGTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1036
DB 861 TTGTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 920
QY 1037 TATGAAAGAAATCCCTTACCTATCAAGAAACAAACACTCTGTGTT 1083
DB 921 TACGAAAGAAATCCCTTACCTATCAAGAAACAAACACTCTGTGTT 967

RESULT 14
BX370641 908 bp mRNA linear EST 08-MAY-2003
LOCUS BX370641 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL004YPO3 5-PRIME, mRNA sequence.
ACCESSION BX370641
VERSION BX370641.1 GI:30459737
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 908)
AUTHORS Li, W.B., Gruber, C., Jessee, U. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG053ZH03_CS05016_1&cluster=7238.f.
Contact: Feng Liang Email: filiang@liferetech.com URL:
http://fulllength.invitrogen.com/invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CS0BAG053ZH03_CS05016_1.

FEATURES

Location/Qualifiers

source

1. .908
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODL004P03"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match 59.5%; Score 645.4; DB 13; Length 908;
Best Local Similarity 98.2%; Pred. No. 1e-130;
Matches 663; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

ORIGIN

410 GATGGTATGAACCTTTCCAACTTGGCACTGTATATTTCCCATCTCCGACCTGAATG 469
1 GATGGTATGAACCTTTCCAACTTGGCACTGTATATTTCCCATCTCCGACCTGAATG 59
470 GATGCCCCCTTCTGGTATATCAAGGCGCTGCTTTTGGAGGAATTGATGATGT 529
60 GATGCCCCCTTCTGGTATATCAAGGCGCTGCTTTTGGAGGAATTGATGATGT 119
530 CACTGGAGGAAATGGGACATTAAGTCAAGTATATCAAGGAAACATGTTCAAC 589
120 CACTGGAGGAAATGGGACATTAAGTCAAGTATATCAAGGAAACATGTTCAAC 179
590 CAAATGGCAAGTGGGTGAACAGACATGAATGAACAGAAATTATATGAGCATGGAT 649
180 CAAATGGCAAGTGGGTGAACAGACATGAATGAACAGAAATTATATGAGCATGGAT 239
650 GTAAGAAGCAAGCCGAAAGAGGAGAGAGATGTTGATCTCAAGACTGTTCCAAA 709
240 GTAAGAAGCAAGCCGAAAGAGGAGAGAGATGTTGATCTCAAGACTGTTCCAAA 299
710 TTTGTTTAAAGACCTTTAACAAGTTGCTGAATTTGAGACAGAGTTCAAGAACATGAA 769
300 TTTGTTTAAAGACCTTTAACAAGTTGCTGAATTTGAGACAGAGTTCAAGAACATGAA 359
770 ACCAATATACAAAGATATTTCTTCACTGAGGAAACCTACTATATCGGAAATGAAAC 829
360 ACCAATATACAAAGATATTTCTTCACTGAGGAAACCTACTATATCGGAAATGAAAC 419
830 TCTGTTTGGGCAACAGAAACAGACTCTGTTGTTAGCCATAAAGATTTATAC 889
420 TCTGTTTGGGCAACAGAAACAGACTCTGTTGTTAGCCATAAAGATTTATAC 479
890 CCGTTCACCAACATTTGGCCACTAAGAAATTTCTGTAGTCTCTTGAATTTTGAAT 549
480 CCGTTCACCAACATTTGGCCACTAAGAAATTTCTGTAGTCTCTTGAATTTTGAAT 539
950 GCAAGTATGTCACAAACAGTTCTATTTGTTTATATTTGAATATGTTTAACT 1009
540 GCAAGTATGTCACAAACAGTTCTATTTGTTTATATTTGAATATGTTTAACT 599
1010 ATGAATTCCTCTTATTAATTAATATGAAGAAATCCCTTACTCTATGAAACAAA 1069
600 ATGAATTCCTCTTATTAATTAATATGAAGAAATCCCTTACTCTATGAAACAAA 559
QY 1070 ACACTCTCTGTTTA 1084
DB 660 ACACTCTCTGTTTA 674

RESULT 15
CB215456 641 bp mRNA linear EST 06-FEB-2003
LOCUS CB215456
DEFINITION NISC np06c07.y1 NICHD_HS_Utl1 Homo sapiens cDNA clone IMAGE:5937013
5', mRNA sequence.

ACCESSION CB215456
VERSION CB215456.1 GI:28263648
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 641)
NCT-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: ccgaps-rc@mail.nih.gov
CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
info@image.lnl.gov
Plate: LLM13163 row: E column: 14
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1. .641
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5937013"
/sex="female"
/tissue_type="normal endometrium, late proliferative
phase, cycle day 13"
/lab_host="DH10B (TI-resistant)"
/clone_lib="NICHD_HS_Utl1"
/note="Organ: uterus; Vector: pCMV-Sport6.1.cdb (Resgen,
Invitrogen Corporation); Site 1: NotI; Site 2: EcoRV;
Cloned unidirectionally from microquantity amounts of mRNA
from normal endometrial tissue (late proliferative phase,
cycle day 13). Average insert size 1.9 kb. Library
constructed by Resgen (Invitrogen Corporation)."

FEATURES

source

ORIGIN

Query Match 59.1%; Score 641; DB 14; Length 641;
Best Local Similarity 100.0%; Pred. No. 9.3e-130;
Matches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

96 CGCTGGCGTCTGTTGGCTCGCGGTGTTCCGGCTGTCCTCGGCAATCCCT 155
1 CGCTGGCGTCTGTTGGCTCGCGGTGTTCCGGCTGTCCTCGGCAATCCCT 60
156 CCGGCGGCAACAGCGGCGGCTCAAGCGCTTGACTTCGCTCCAAAACCTGATCCTT 215
61 CCGGCGGCAACAGCGGCGGCTCAAGCGCTTGACTTCGCTCCAAAACCTGATCCTT 120
216 ATTGTCAAGCTAAGTATCTTCTGTCCAACTGGCTCACTATCCAGTTATGAGAGGTG 275
121 ATTGTCAAGCTAAGTATCTTCTGTCCAACTGGCTCACTATCCAGTTATGAGAGGTG 180
276 ATGATGACATGAAGTTTTCATTAACAAGCCCGAGTATGGAATTTAATGAGAAC 335
181 ATGATGACATGAAGTTTTCATTAACAAGCCCGAGTATGGAATTTAATGAGAAC 240
336 TCGTGGAGACCTGAAATTAATGACATGATGATGATGATGATGATGATGATGATGAT 395
241 TCGTGGAGACCTGAAATTAATGACATGATGATGATGATGATGATGATGATGATGAT 300
396 AGAATACACAATGGAATGATGAACTTTTCCAACTGGCACTGTACATTTCCCATC 455
301 AGAATACACAATGGAATGATGAACTTTTCCAACTGGCACTGTACATTTCCCATC 360
456 TCGGACCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 515
361 TCGGACCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420

```
Qy 516 GAATGATGATGTTCACTGGAGGAAATGGACATTAGTTCAGTACACTATATCAG 575
Db 421 GAATGATGATGTTCACTGGAGGAAATGGACATTAGTTCAGTACACTATATCAG 480
Qy 576 GAAACATGTTCAACCAATGGCAAAGTGGGTGAACAGACAAATGAAACAGGAAATTTATT 635
Db 481 GAAACATGTTCAACCAATGGCAAAGTGGGTGAACAGACAAATGAAACAGGAAATTTATT 540
Qy 636 ATGAGACATGGAATGTAAAGCCAGCCCAAGAAAGGGGAGAGACATGTTGATTCCCT 695
Db 541 ATGAGACATGGAATGTAAAGCCAGCCCAAGAAAGGGGAGAGACATGTTGATTCCCT 600
Qy 696 ACGACTGTTCCAAATTGTGTAAAGACCTTTAACAGTTG 736
Db 601 ACGACTGTTCCAAATTGTGTAAAGACCTTTAACAGTTG 641
```

Search completed: Apr:11 22, 2004, 04:39:45
Job time : 3276.24 secs

[illegible]

724 TTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCACTATACAGAAATATTTCTT 793
 726 TTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCACTATACAGAAATATTTCTT 795
 728 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 853
 729 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 730 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 731 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 732 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 733 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 734 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 735 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 736 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 737 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 738 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 739 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 740 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 741 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 742 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 743 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 744 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 745 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 746 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 747 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 748 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 749 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 750 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 751 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 752 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 753 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 754 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 755 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 756 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 757 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 758 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 759 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 760 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 761 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 762 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 763 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 764 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 765 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 766 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 767 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 768 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 769 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 770 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 771 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 772 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 773 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 774 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 775 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 776 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 777 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 778 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 779 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 780 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 781 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 782 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 783 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 784 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 785 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 786 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 787 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 788 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 789 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 790 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 791 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 792 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 793 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 794 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 795 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 796 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 797 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 798 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 799 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855
 800 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 855

RESULT 4

US-09-397-945-26
 Sequence 26, Application US/09397945
 Publication No. US20030065139A1
 GENERAL INFORMATION:
 APPLICANT: Human Genome Sciences, Inc. et al.
 TITLE OF INVENTION: 95 Human secreted proteins
 FILE REFERENCE: P2027P1
 CURRENT APPLICATION NUMBER: US/09/397,945
 CURRENT FILING DATE: 1999-09-17
 PRIOR APPLICATION NUMBER: PCT/US99/05804
 PRIOR FILING DATE: 1999-03-18
 PRIOR APPLICATION NUMBER: 60/078,566
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,576
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,573
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,574
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,579
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/080,314
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/080,312
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/078,578
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,581
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,577
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,563
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/080,313
 PRIOR FILING DATE: 1998-04-01
 NUMBER OF SEQ ID NOS: 470
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 26
 LENGTH: 1751
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (1520)
 OTHER INFORMATION: n equals a,t,g, or c
 NAME/KEY: SITE
 LOCATION: (1557)

OTHER INFORMATION: n equals a,t,g, or c
 NAME/KEY: SITE
 LOCATION: (1689)
 OTHER INFORMATION: n equals a,t,g, or c
 NAME/KEY: SITE
 LOCATION: (1729)
 OTHER INFORMATION: n equals a,t,g, or c
 NAME/KEY: SITE
 LOCATION: (1735)
 OTHER INFORMATION: n equals a,t,g, or c
 NAME/KEY: SITE
 LOCATION: (1741)
 OTHER INFORMATION: n equals a,t,g, or c
 US-09-397-945-26

Query Match 98.4%; Score 1067; DB 13; Length 1751;
 Best Local Similarity 99.6%; Pred. No. 3.2e-278;
 Matches 1067; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

14 GGGGAGAGGTAAGACACGGGCAAGGCGCGGAGATGCGGCGGCGGCGGCGGCGGCGG 73
 16 GCGGAGAGGTAAGACACGGGCAAGGCGCGGAGATGCGGCGGCGGCGGCGGCGGCGG 75
 74 GAGGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 133
 76 GAGGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 135
 134 TCCCGGCTCTGCGGCGGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 193
 136 TCCCGGCTCTGCGGCGGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 195
 194 TTCGCTCCAAAACCGATCCCTATTTGCAAGCTAAGTAACTTCTGTCGCACTGCTCA 253
 196 TTCGCTCCAAAACCGATCCCTATTTGCAAGCTAAGTAACTTCTGTCGCACTGCTCA 255
 254 CCTATCCAGTATGAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 313
 256 CCTATCCAGTATGAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 315
 314 TGGGAAATTAATATGAGAGACCTCTGCGGACCTTGAATTAATGATGATGATGATGAT 373
 316 TGGGAAATTAATATGAGAGACCTCTGCGGACCTTGAATTAATGATGATGATGATGAT 375
 374 TACAGAAAGTAACTTAACTGAGAAAGTAACTGAGAAAGTAACTGAGAAAGTAACTT 433
 376 TACAGAAAGTAACTTAACTGAGAAAGTAACTGAGAAAGTAACTGAGAAAGTAACTT 435
 434 GGGAACTGTAATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGAT 493
 436 GGGAACTGTAATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGAT 495
 494 GGGGCTGCTGCTTCTTCTTCTGAGGAAATGATGATGATGATGATGATGATGATGAT 553
 496 GGGGCTGCTGCTTCTTCTTCTGAGGAAATGATGATGATGATGATGATGATGATGAT 555
 554 GTTCAAGTGAATCTTATGAGAAACATGTTCAACCAATGAGAAAGTGGTGAACAG 613
 556 GTTCAAGTGAATCTTATGAGAAACATGTTCAACCAATGAGAAAGTGGTGAACAG 615
 614 GACATGAACAGAAATTTATATGAGACATGAAATGTAAGAAAGCCAGAAAGGGG 673
 616 GACATGAACAGAAATTTATATGAGACATGAAATGTAAGAAAGCCAGAAAGGGG 675
 674 GCGAGACATGTTGATTTCTGAGGAAATGATGATGATGATGATGATGATGATGATGAT 733
 676 GCGAGACATGTTGATTTCTGAGGAAATGATGATGATGATGATGATGATGATGATGAT 735
 734 TTGGCTGAATTTGGAGAGAGTTCAAGAACTAGAAACCACTATACAGAAATATTTCTT 793
 736 TTGGCTGAATTTGGAGAGAGTTCAAGAACTAGAAACCACTATACAGAAATATTTCTT 795
 794 TACAGTGGAGAACTTACTTATCTGGGAAATGAAACATCTGTTTGGCCCAACAGAAAC 853